(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
10 October 2002 (10.10.2002)

PCT

(10) International Publication Number
WO 02/078420 A2

(51) International Patent Classification: Not classified

(21) International Application Number: PCT/US02/09809

(22) International Filing Date: 29 March 2002 (29.03.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/280,387	30 March 2001 (30.03.2001)	US
60/282,335	5 April 2001 (05.04.2001)	US
60/283,663	13 April 2001 (13.04.2001)	US
60/285,484	19 April 2001 (19.04.2001)	US
60/350,702	18 January 2002 (18.01.2002)	US
60/351,749	25 January 2002 (25.01.2002)	US
60/351,749	25 January 2002 (25.01.2002)	US

(71) Applicant *(for all designated States except US)*: **INCYTE GENOMICS, INC.** [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **LU, Dyung, Aina, M.** [US/US]; 233 Coy Drive, San Jose, CA 95123 (US). **ARVIZU, Chandra, S.** [US/US]; 1706 Morocco Drive, San Jose, CA 95125 (US). **GANDHI, Ameena, R.** [US/US]; 705 5th Avenue, San Francisco, CA 94118 (US). **HAFALIA, April, J., A.** [US/US]; 2227 Calle de Primavera, Santa Clara, CA 95054 (US). **DING, Li** [CN/US]; 3353 Alma Street # 146, Palo Alto, CA 94306 (US). **LU, Yan** [CN/US]; 3885 Corrina Way, Palo Alto, CA 94303 (US). **RAMKUMAR, Jayalaxmi** [IN/US]; 34359 Maybird Circle, Fremont, CA 94555 (US). **SWARNAKAR, Anita** [CA/US]; 8 Locksley Avenue #5D, San Francisco, CA 94122 (US). **TANG, Y., Tom** [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). **YUE, Henry** [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). **TRAN, Bao** [US/US]; 750 Salberg Avenue, Santa Clara, CA 95051 (US). **LEE, Soo, Y.** [KR/US]; 40 Westdale Avenue, Daly City, CA 94015 (US). **WARREN, Bridget, A.** [US/US]; 2250 Homestead Court #2, Los Altos, CA 94024 (US). **NGUYEN, Danniel, B.** [US/US]; 1403 Ridgewood Drive, San Jose, CA 95118 (US). **THANGAVELU, Kavitha** [IN/US]; 1950 Montecito Avenue #23, Mountain View, CA 94043 (US). **YAO, Monique, G.** [US/US]; 1189 Woodgate Drive, Carmel, IN 46033 (US). **ELLIOTT, Vicki, S.** [US/US]; 3770 Polton Place Way, San Jose, CA 95121 (US). **BAUGHN, Mariah, R.** [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). **EMERLING, Brooke, M.** [US/US]; 1735 Woodland Avenue # 71, Palo Alto, CA 94303 (US). **LAL, Preeti, G.** [US/US]; P.O. Box 5142, Santa Clara, CA 95056 (US). **GIETZEN, Kimberly, J.** [US/US]; 691 Los Huecos Drive, San Jose, CA 95123 (US). **BECHA, Shanya, D.** [US/US]; 21062 Gary Drive # 117, Castro Valley, CA 94546 (US). **MARQUIS, Joseph, P.** [US/US]; 4428 Lazy Lane, San Jose, CA 95135 (US). **KABLE, Amy, E.** [US/US]; 2345 Polk Street #4, San Francisco, CA 94109 (US).

(74) Agents: **HAMLET-COX, Diana** et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*
— *entirely in electronic form (except for this front page) and available upon request from the International Bureau*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: MOLECULES FOR DISEASE DETECTION AND TREATMENT

(57) Abstract: The invention provides human molecules for disease detection and treatment (MDDT) and polynucleotides which identify and encode MDDT. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of MDDT.

WO 02/078420 A2

MOLECULES FOR DISEASE DETECTION AND TREATMENT

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of molecules for disease detection and treatment and to the use of these sequences in the diagnosis, treatment, and prevention of cell proliferative, autoimmune/inflammatory, developmental, and neurological disorders, and infections, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of molecules for disease detection and treatment.

BACKGROUND OF THE INVENTION

It is estimated that only 2% of mammalian DNA encodes proteins, and only a small fraction of the genes that encode proteins are actually expressed in a particular cell at any time. The various types of cells in a multicellular organism differ dramatically both in structure and function, and the identity of a particular cell is conferred by its unique pattern of gene expression. In addition, different cell types express overlapping but distinctive sets of genes throughout development. Cell growth and proliferation, cell differentiation, the immune response, apoptosis, and other processes that contribute to organismal development and survival are governed by regulation of gene expression. Appropriate gene regulation also ensures that cells function efficiently by expressing only those genes whose functions are required at a given time. Factors that influence gene expression include extracellular signals that mediate cell-cell communication and coordinate the activities of different cell types. Gene expression is regulated at the level of DNA and RNA transcription, and at the level of mRNA translation.

Aberrant expression or mutations in genes and their products may cause, or increase susceptibility to, a variety of human diseases such as cancer and other cell proliferative disorders. The identification of these genes and their products is the basis of an ever-expanding effort to find markers for early detection of diseases and targets for their prevention and treatment. For example, cancer represents a type of cell proliferative disorder that affects nearly every tissue in the body. The development of cancer, or oncogenesis, is often correlated with the conversion of a normal gene into a cancer-causing gene, or oncogene, through abnormal expression or mutation. Oncoproteins, the products of oncogenes, include a variety of molecules that influence cell proliferation, such as growth factors, growth factor receptors, intracellular signal transducers, nuclear transcription factors, and cell-cycle control proteins. In contrast, tumor-suppressor genes are involved in inhibiting cell proliferation. Mutations which reduce or abrogate the function of tumor-suppressor genes result in

aberrant cell proliferation and cancer. Thus a wide variety of genes and their products have been found that are associated with cell proliferative disorders such as cancer, but many more may exist that are yet to be discovered.

DNA-based arrays can provide an efficient, high-throughput method to examine gene expression and genetic variability. For example, SNPs, or single nucleotide polymorphisms, are the most common type of human genetic variation. DNA-based arrays can dramatically accelerate the discovery of SNPs in hundreds and even thousands of genes. Likewise, such arrays can be used for SNP genotyping in which DNA samples from individuals or populations are assayed for the presence of selected SNPs. These approaches will ultimately lead to the systematic identification of all genetic variations in the human genome and the correlation of certain genetic variations with disease susceptibility, responsiveness to drug treatments, and other medically relevant information. (See, for example, Wang, D.G. et al. (1998) Science 280:1077-1082.)

DNA-based array technology is especially important for the rapid analysis of global gene expression patterns. For example, genetic predisposition, disease, or therapeutic treatment may directly or indirectly affect the expression of a large number of genes in a given tissue. In this case, it is useful to develop a profile, or transcript image, of all the genes that are expressed and the levels at which they are expressed in that particular tissue. A profile generated from an individual or population affected with a certain disease or undergoing a particular therapy may be compared with a profile generated from a control individual or population. Such analysis does not require knowledge of gene function, as the expression profiles can be subjected to mathematical analyses which simply treat each gene as a marker. Furthermore, gene expression profiles may help dissect biological pathways by identifying all the genes expressed, for example, at a certain developmental stage, in a particular tissue, or in response to disease or treatment. (See, for example, Lander, E.S. et al. (1996) Science 274:536-539.)

Certain genes are known to be associated with diseases because of their chromosomal location, such as the genes in the myotonic dystrophy (DM) regions of mouse and human. The mutation underlying DM has been localized to a gene encoding the DM-kinase protein, but another active gene, DMR-N9, is in close proximity to the DM-kinase gene (Jansen, G. et al. (1992) Nat. Genet. 1:261-266). DMR-N9 encodes a 650 amino acid protein that contains WD repeats, motifs found in cell signaling proteins. DMR-N9 is expressed in all neural tissues and in the testis, suggesting a role for DMR-N9 in the manifestation of mental and testicular symptoms in severe cases of DM (Jansen, G. et al. (1995) Hum. Mol. Genet. 4:843-852).

Other genes are identified based upon their expression patterns or association with disease

syndromes. For example, autoantibodies to subcellular organelles are found in patients with systemic rheumatic diseases. A recently identified protein, golgin-67, belongs to a family of Golgi autoantigens having alpha-helical coiled-coil domains (Eystathioy, T. et al. (2000) J. Autoimmun. 14:179-187). The Stac gene was identified as a brain specific, developmentally regulated gene. The Stac protein contains an SH3 domain, and is thought to be involved in neuron-specific signal transduction (Suzuki, H. et al. (1996) Biochem. Biophys. Res. Commun. 229:902-909).

<u>Structural and Cytoskeleton-Associated Proteins</u>

The cytoskeleton is a cytoplasmic network of protein fibers that mediate cell shape, structure, and movement. The cytoskeleton supports the cell membrane and forms tracks along which organelles and other elements move in the cytosol. The cytoskeleton is a dynamic structure that allows cells to adopt various shapes and to carry out directed movements. Major cytoskeletal fibers include the microtubules, the microfilaments, and the intermediate filaments. Motor proteins, including myosin, dynein, and kinesin, drive movement of or along the fibers. The motor protein dynamin drives the formation of membrane vesicles. Accessory or associated proteins modify the structure or activity of the fibers while cytoskeletal membrane anchors connect the fibers to the cell membrane.

Microtubules and Associated Proteins

<u>Tubulins</u>

Microtubules, cytoskeletal fibers with a diameter of about 24 nm, have multiple roles in the cell. Bundles of microtubules form cilia and flagella, which are whip-like extensions of the cell membrane that are necessary for sweeping materials across an epithelium and for swimming of sperm, respectively. Marginal bands of microtubules in red blood cells and platelets are important for these cells' pliability. Organelles, membrane vesicles, and proteins are transported in the cell along tracks of microtubules. For example, microtubules run through nerve cell axons, allowing bi-directional transport of materials and membrane vesicles between the cell body and the nerve terminal. Failure to supply the nerve terminal with these vesicles blocks the transmission of neural signals. Microtubules are also critical to chromosomal movement during cell division. Both stable and short-lived populations of microtubules exist in the cell.

Microtubules are polymers of GTP-binding tubulin protein subunits. Each subunit is a heterodimer of α- and β- tubulin, multiple isoforms of which exist. The hydrolysis of GTP is linked to the addition of tubulin subunits at the end of a microtubule. The subunits interact head to tail to form protofilaments; the protofilaments interact side to side to form a microtubule. A microtubule is polarized, one end ringed with α-tubulin and the other with β-tubulin, and the two ends differ in their rates of assembly. Generally, each microtubule is composed of 13 protofilaments although 11 or 15

protofilament-microtubules are sometimes found. Cilia and flagella contain doublet microtubules. Microtubules grow from specialized structures known as centrosomes or microtubule-organizing centers (MTOCs). MTOCs may contain one or two centrioles, which are pinwheel arrays of triplet microtubules. The basal body, the organizing center located at the base of a cilium or flagellum, contains one centriole. Gamma tubulin present in the MTOC is important for nucleating the polymerization of α- and β- tubulin heterodimers but does not polymerize into microtubules.

Microtubule-Associated Proteins

Microtubule-associated proteins (MAPs) have roles in the assembly and stabilization of microtubules. One major family of MAPs, assembly MAPs, can be identified in neurons as well as non-neuronal cells. Assembly MAPs are responsible for cross-linking microtubules in the cytosol. These MAPs are organized into two domains: a basic microtubule-binding domain and an acidic projection domain. The projection domain is the binding site for membranes, intermediate filaments, or other microtubules. Based on sequence analysis, assembly MAPs can be further grouped into two types: Type I and Type II. Type I MAPs, which include MAP1A and MAP1B, are large, filamentous molecules that co-purify with microtubules and are abundantly expressed in brain and testes. Type I MAPs contain several repeats of a positively-charged amino acid sequence motif that binds and neutralizes negatively charged tubulin, leading to stabilization of microtubules. MAP1A and MAP1B are each derived from a single precursor polypeptide that is subsequently proteolytically processed to generate one heavy chain and one light chain.

Another light chain, LC3, is a 16.4 kDa molecule that binds MAP1A, MAP1B, and microtubules. It is suggested that LC3 is synthesized from a source other than the MAP1A or MAP1B transcripts, and that the expression of LC3 may be important in regulating the microtubule binding activity of MAP1A and MAP1B during cell proliferation (Mann, S.S. et al. (1994) J. Biol. Chem. 269:11492-11497).

Type II MAPs, which include MAP2a, MAP2b, MAP2c, MAP4, and Tau, are characterized by three to four copies of an 18-residue sequence in the microtubule-binding domain. MAP2a, MAP2b, and MAP2c are found only in dendrites, MAP4 is found in non-neuronal cells, and Tau is found in axons and dendrites of nerve cells. Alternative splicing of the Tau mRNA leads to the existence of multiple forms of Tau protein. Tau phosphorylation is altered in neurodegenerative disorders such as Alzheimer's disease, Pick's disease, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia and Parkinsonism linked to chromosome 17. The altered Tau phosphorylation leads to a collapse of the microtubule network and the formation of intraneuronal Tau aggregates (Spillantini, M.G. and M. Goedert (1998) Trends Neurosci. 21:428-433).

Another microtubule associated protein, STOP (stable tubule only polypeptide), is a calmodulin-regulated protein that regulates stability (Denarier, E. et al. (1998) Biochem. Biophys. Res. Commun. 24:791-796). In order for neurons to maintain conductive connections over great distances, they rely upon axodendritic extensions, which in turn are supported by microtubules. STOP proteins function to stabilize the microtubular network. STOP proteins are associated with axonal microtubules, and are also abundant in neurons (Guillaud, L. et al. (1998) J. Cell Biol. 142:167-179). STOP proteins are necessary for normal neurite formation, and have been observed to stabilize microtubules, *in vitro*, against cold-, calcium-, or drug-induced dissassembly (Margolis, R.L. et al. (1990) EMBO 9:4095-502).

Microfilaments and Associated Proteins

Actins

Microfilaments, cytoskeletal filaments with a diameter of about 7-9 nm, are vital to cell locomotion, cell shape, cell adhesion, cell division, and muscle contraction. Assembly and disassembly of the microfilaments allow cells to change their morphology. Microfilaments are the polymerized form of actin, the most abundant intracellular protein in the eukaryotic cell. Human cells contain six isoforms of actin. The three α-actins are found in different kinds of muscle, nonmuscle β-actin and nonmuscle γ-actin are found in nonmuscle cells, and another γ-actin is found in intestinal smooth muscle cells. G-actin, the monomeric form of actin, polymerizes into polarized, helical F-actin filaments, accompanied by the hydrolysis of ATP to ADP. Actin filaments associate to form bundles and networks, providing a framework to support the plasma membrane and determine cell shape. These bundles and networks are connected to the cell membrane. In muscle cells, thin filaments containing actin slide past thick filaments containing the motor protein myosin during contraction. A family of actin-related proteins exist that are not part of the actin cytoskeleton, but rather associate with microtubules and dynein.

Actin-Associated Proteins

Actin-associated proteins have roles in cross-linking, severing, and stabilization of actin filaments and in sequestering actin monomers. Several of the actin-associated proteins have multiple functions. Bundles and networks of actin filaments are held together by actin cross-linking proteins. These proteins have two actin-binding sites, one for each filament. Short cross-linking proteins promote bundle formation while longer, more flexible cross-linking proteins promote network formation. Actin-interacting proteins (AIPs) participate in the regulation of actin filament organization. Other actin-associated proteins such as TARA, a novel F-actin binding protein, function in a similar capacity by regulating actin cytoskeletal organization. Calmodulin-like calcium-binding

domains in actin cross-linking proteins allow calcium regulation of cross-linking. Group I cross-linking proteins have unique actin-binding domains and include the 30 kD protein, EF-1a, fascin, and scruin. Group II cross-linking proteins have a 7,000-MW actin-binding domain and include villin and dematin. Group III cross-linking proteins have pairs of a 26,000-MW actin-binding domain and include fimbrin, spectrin, dystrophin, ABP 120, and filamin.

Severing proteins regulate the length of actin filaments by breaking them into short pieces or by blocking their ends. Severing proteins include gCAP39, severin (fragmin), gelsolin, and villin. Capping proteins can cap the ends of actin filaments, but cannot break filaments. Capping proteins include CapZ and tropomodulin. The proteins thymosin and profilin sequester actin monomers in the cytosol, allowing a pool of unpolymerized actin to exist. The actin-associated proteins tropomyosin, troponin, and caldesmon regulate muscle contraction in response to calcium.

Microtubule and actin filament networks cooperate in processes such as vesicle and organelle transport, cleavage furrow placement, directed cell migration, spindle rotation, and nuclear migration. Microtubules and actin may coordinate to transport vesicles, organelles, and cell fate determinants, or transport may involve targeting and capture of microtubule ends at cortical actin sites. These cytoskeletal systems may be bridged by myosin-kinesin complexes, myosin-CLIP170 complexes, formin-homology (FH) proteins, dynein, the dynactin complex, Kar9p, coronin, ERM proteins, and kelch repeat-containing proteins (for a review, see Goode, B.L. et al. (2000) Curr. Opin. Cell Biol. 12:63-71). The kelch repeat is a motif originally observed in the kelch protein, which is involved in formation of cytoplasmic bridges called ring canals. A variety of mammalian and other kelch family proteins have been identified. The kelch repeat domain is believed to mediate interaction with actin (Robinson, D.N. and L. Cooley (1997) J. Cell Biol. 138:799-810).

ADF/cofilins are a family of conserved 15-18 kDa actin-binding proteins that play a role in cytokinesis, endocytosis, and in development of embryonic tissues, as well as in tissue regeneration and in pathologies such as ischemia, oxidative or osmotic stress. LIM kinase 1 downregulates ADF (Carlier, M.F. et al. (1999) J. Biol. Chem. 274:33827-33830).

LIM is an acronym of three transcription factors, Lin-11, Isl-1, and Mec-3, in which the motif was first identified. The LIM domain is a double zinc-finger motif that mediates the protein-protein interactions of transcription factors, signaling, and cytoskeleton-associated proteins (Roof, D.J. et al. (1997) J. Cell Biol. 138:575-588). These proteins are distributed in the nucleus, cytoplasm, or both (Brown, S. et al. (1999) J. Biol. Chem. 274:27083-27091). Recently, ALP (actinin-associated LIM protein) has been shown to bind alpha-actinin-2 (Bouju, S. et al. (1999) Neuromuscul. Disord. 9:3-10).

The Frabin protein is another example of an actin-filament binding protein (Obaishi, H. et al.

(1998) J. Biol. Chem. 273:18697-18700). Frabin (FGD1-related F-actin-binding protein) possesses one actin-filament binding (FAB) domain, one Dbl homology (DH) domain, two pleckstrin homology (PH) domains, and a single cysteine-rich FYVE (Fab1p, YOTB, Vac1p, and EEA1 (early endosomal antigen 1)) domain. Frabin has shown GDP/GTP exchange activity for Cdc42 small G protein (Cdc42), and indirectly induces activation of Rac small G protein (Rac) in intact cells. Through the activation of Cdc42 and Rac, Frabin is able to induce formation of both filopodia- and lamellipodia-like processes (Ono, Y. et al. (2000) Oncogene 19:3050-3058). The Rho family small GTP-binding proteins are important regulators of actin-dependent cell functions including cell shape change, adhesion, and motility. The Rho family consists of three major subfamilies: Cdc42, Rac, and Rho. Rho family members cycle between GDP-bound inactive and GTP-bound active forms by means of a GDP/GTP exchange factor (GEF) (Umikawa, M. et al. (1999) J. Biol. Chem. 274:25197-25200). The Rho GEF family is crucial for microfilament organization.

Intermediate Filaments and Associated Proteins

Intermediate filaments (IFs) are cytoskeletal fibers with a diameter of about 10 nm, intermediate between that of microfilaments and microtubules. IFs serve structural roles in the cell, reinforcing cells and organizing cells into tissues. IFs are particularly abundant in epidermal cells and in neurons. IFs are extremely stable, and, in contrast to microfilaments and microtubules, do not function in cell motility.

Five types of IF proteins are known in mammals. Type I and Type II proteins are the acidic and basic keratins, respectively. Heterodimers of the acidic and basic keratins are the building blocks of keratin IFs. Keratins are abundant in soft epithelia such as skin and cornea, hard epithelia such as nails and hair, and in epithelia that line internal body cavities. Mutations in keratin genes lead to epithelial diseases including epidermolysis bullosa simplex, bullous congenital ichthyosiform erythroderma (epidermolytic hyperkeratosis), non-epidermolytic and epidermolytic palmoplantar keratoderma, ichthyosis bullosa of Siemens, pachyonychia congenita, and white sponge nevus. Some of these diseases result in severe skin blistering. (See, e.g., Wawersik, M. et al. (1997) J. Biol. Chem. 272:32557-32565; and Corden L.D. and W.H. McLean (1996) Exp. Dermatol. 5:297-307.)

Type III IF proteins include desmin, glial fibrillary acidic protein, vimentin, and peripherin. Desmin filaments in muscle cells link myofibrils into bundles and stabilize sarcomeres in contracting muscle. Glial fibrillary acidic protein filaments are found in the glial cells that surround neurons and astrocytes. Vimentin filaments are found in blood vessel endothelial cells, some epithelial cells, and mesenchymal cells such as fibroblasts, and are commonly associated with microtubules. Vimentin filaments may have roles in keeping the nucleus and other organelles in place in the cell. Type IV IFs

include the neurofilaments and nestin. Neurofilaments, composed of three polypeptides NF-L, NF-M, and NF-H, are frequently associated with microtubules in axons. Neurofilaments are responsible for the radial growth and diameter of an axon, and ultimately for the speed of nerve impulse transmission. Changes in phosphorylation and metabolism of neurofilaments are observed in neurodegenerative diseases including amyotrophic lateral sclerosis, Parkinson's disease, and Alzheimer's disease (Julien, J.P. and W.E. Mushynski (1998) Prog. Nucleic Acid Res. Mol. Biol. 61:1-23). Type V IFs, the lamins, are found in the nucleus where they support the nuclear membrane.

IFs have a central α-helical rod region interrupted by short nonhelical linker segments. The rod region is bracketed, in most cases, by non-helical head and tail domains. The rod regions of intermediate filament proteins associate to form a coiled-coil dimer. A highly ordered assembly process leads from the dimers to the IFs. Neither ATP nor GTP is needed for IF assembly, unlike that of microfilaments and microtubules.

IF-associated proteins (IFAPs) mediate the interactions of IFs with one another and with other cell structures. IFAPs cross-link IFs into a bundle, into a network, or to the plasma membrane, and may cross-link IFs to the microfilament and microtubule cytoskeleton. Microtubules and IFs are particularly closely associated. IFAPs include BPAG1, plakoglobin, desmoplakin I, desmoplakin II, plectin, ankyrin, filaggrin, and lamin B receptor.

Cytoskeletal-Membrane Anchors

Cytoskeletal fibers are attached to the plasma membrane by specific proteins. These attachments are important for maintaining cell shape and for muscle contraction. In erythrocytes, the spectrin-actin cytoskeleton is attached to the cell membrane by three proteins, band 4.1, ankyrin, and adducin. Defects in this attachment result in abnormally shaped cells which are more rapidly degraded by the spleen, leading to anemia. In platelets, the spectrin-actin cytoskeleton is also linked to the membrane by ankyrin; a second actin network is anchored to the membrane by filamin. In muscle cells the protein dystrophin links actin filaments to the plasma membrane; mutations in the dystrophin gene lead to Duchenne muscular dystrophy.

Focal adhesions

Focal adhesions are specialized structures in the plasma membrane involved in the adhesion of a cell to a substrate, such as the extracellular matrix. Focal adhesions form the connection between an extracellular substrate and the cytoskeleton, and affect such functions as cell shape, cell motility and cell proliferation. Transmembrane integrin molecules form the basis of focal adhesions. Upon ligand binding, integrins cluster in the plane of the plasma membrane. Cytoskeletal linker proteins such as the actin binding proteins a-actinin, talin, tensin, vinculin, paxillin, and filamin are recruited to the

clustering site. Key regulatory proteins, such as Rho and Ras family proteins, focal adhesion kinase, and Src family members are also recruited. These events lead to the reorganization of actin filaments and the formation of stress fibers. These intracellular rearrangements promote further integrin-ECM interactions and integrin clustering. Thus, integrins mediate aggregation of protein complexes on both the cytosolic and extracellular faces of the plasma membrane, leading to the assembly of the focal adhesion. Many signal transduction responses are mediated via various adhesion complex proteins, including Src, FAK, paxillin, and tensin. (For a review, see Yamada, K.M. and B. Geiger, (1997) Curr. Opin. Cell Biol. 9:76-85.)

IFs are also attached to membranes by cytoskeletal-membrane anchors. The nuclear lamina is attached to the inner surface of the nuclear membrane by the lamin B receptor. Vimentin IFs are attached to the plasma membrane by ankyrin and plectin. Desmosome and hemidesmosome membrane junctions hold together epithelial cells of organs and skin. These membrane junctions allow shear forces to be distributed across the entire epithelial cell layer, thus providing strength and rigidity to the epithelium. IFs in epithelial cells are attached to the desmosome by plakoglobin and desmoplakins. The proteins that link IFs to hemidesmosomes are not known. Desmin IFs surround the sarcomere in muscle and are linked to the plasma membrane by paranemin, synemin, and ankyrin.

The protein components of tight junctions include ZO-1 and ZO-2 (zona occludens), cytoplasmic proteins associated with the plasma membrane at tight junctions. ZO-1 is a PDZ domain-containing protein which associates with spectrin and thus may link tight junctions to the actin cytoskeleton. Other cytoplasmic components of tight junctions include cingulin, 7H6 antigen, symplekin, and small rab family GTPases. The first identified component of the tight junction strands, which form the actual junction between cells, was the integral membrane protein occludin, a 65 kD protein with four transmembrane domains. ZO-1 binds to the carboxy-terminal region of occludin and may localize occludin to the tight junction. A recently identified family of proteins, the claudins, are also components of tight junction strands.

Motor Proteins

<u>Myosin-related Motor Proteins</u>

Myosins are actin-activated ATPases, found in eukaryotic cells, that couple hydrolysis of ATP with motion. Myosin provides the motor function for muscle contraction and intracellular movements such as phagocytosis and rearrangement of cell contents during mitotic cell division (cytokinesis). The contractile unit of skeletal muscle, termed the sarcomere, consists of highly ordered arrays of thin actin-containing filaments and thick myosin-containing filaments. Crossbridges form between the thick and thin filaments, and the ATP-dependent movement of myosin heads within the thick filaments pulls

the thin filaments, shortening the sarcomere and thus the muscle fiber.

Myosins are composed of one or two heavy chains and associated light chains. Myosin heavy chains contain an amino-terminal motor or head domain, a neck that is the site of light-chain binding, and a carboxy-terminal tail domain. The tail domains may associate to form an α-helical coiled coil. Conventional myosins, such as those found in muscle tissue, are composed of two myosin heavy-chain subunits, each associated with two light-chain subunits that bind at the neck region and play a regulatory role. Unconventional myosins, believed to function in intracellular motion, may contain either one or two heavy chains and associated light chains. There is evidence for about 25 myosin heavy chain genes in vertebrates, more than half of them unconventional.

Dynein-related Motor Proteins

Dyneins are (-) end-directed motor proteins which act on microtubules. Two classes of dyneins, cytosolic and axonemal, have been identified. Cytosolic dyneins are responsible for translocation of materials along cytoplasmic microtubules, for example, transport from the nerve terminal to the cell body and transport of endocytic vesicles to lysosomes. As well, viruses often take advantage of cytoplasmic dyneins to be transported to the nucleus and establish a successful infection (Sodeik, B. et al. (1997) J. Cell Biol. 136:1007-1021). Virion proteins of herpes simplex virus 1, for example, interact with the cytoplasmic dynein intermediate chain (Ye, G.J. et al. (2000) J. Virol. 74:1355-1363). Cytoplasmic dyneins are also reported to play a role in mitosis. Axonemal dyneins are responsible for the beating of flagella and cilia. Dynein on one microtubule doublet walks along the adjacent microtubule doublet. This sliding force produces bending that causes the flagellum or cilium to beat. Dyneins have a native mass between 1000 and 2000 kDa and contain either two or three force-producing heads driven by the hydrolysis of ATP. The heads are linked via stalks to a basal domain which is composed of a highly variable number of accessory intermediate and light chains. Cytoplasmic dynein is the largest and most complex of the motor proteins.

Kinesin-related Motor Proteins

Kinesins are (+) end-directed motor proteins which act on microtubules. The prototypical kinesin molecule is involved in the transport of membrane-bound vesicles and organelles. This function is particularly important for axonal transport in neurons. Kinesin is also important in all cell types for the transport of vesicles from the Golgi complex to the endoplasmic reticulum. This role is critical for maintaining the identity and functionality of these secretory organelles.

Kinesins define a ubiquitous, conserved family of over 50 proteins that can be classified into at least 8 subfamilies based on primary amino acid sequence, domain structure, velocity of movement, and cellular function. (Reviewed in Moore, J.D. and S.A. Endow (1996) Bioessays 18:207-219; and

Hoyt, A.M. (1994) Curr. Opin. Cell Biol. 6:63-68.) The prototypical kinesin molecule is a heterotetramer comprised of two heavy polypeptide chains (KHCs) and two light polypeptide chains (KLCs). The KHC subunits are typically referred to as "kinesin." KHC is about 1000 amino acids in length, and KLC is about 550 amino acids in length. Two KHCs dimerize to form a rod-shaped molecule with three distinct regions of secondary structure. At one end of the molecule is a globular motor domain that functions in ATP hydrolysis and microtubule binding. Kinesin motor domains are highly conserved and share over 70% identity. Beyond the motor domain is an α-helical coiled-coil region which mediates dimerization. At the other end of the molecule is a fan-shaped tail that associates with molecular cargo. The tail is formed by the interaction of the KHC C-termini with the two KLCs.

Members of the more divergent subfamilies of kinesins are called kinesin-related proteins (KRPs), many of which function during mitosis in eukaryotes (Hoyt, supra). Some KRPs are required for assembly of the mitotic spindle. In vivo and in vitro analyses suggest that these KRPs exert force on microtubules that comprise the mitotic spindle, resulting in the separation of spindle poles. Phosphorylation of KRP is required for this activity. Failure to assemble the mitotic spindle results in abortive mitosis and chromosomal aneuploidy, the latter condition being characteristic of cancer cells. In addition, a unique KRP, centromere protein E, localizes to the kinetochore of human mitotic chromosomes and may play a role in their segregation to opposite spindle poles.

Dynamin-related Motor Proteins

Dynamin is a large GTPase motor protein that functions as a "molecular pinchase," generating a mechanochemical force used to sever membranes. This activity is important in forming clathrin-coated vesicles from coated pits in endocytosis and in the biogenesis of synaptic vesicles in neurons. Binding of dynamin to a membrane leads to dynamin's self-assembly into spirals that may act to constrict a flat membrane surface into a tubule. GTP hydrolysis induces a change in conformation of the dynamin polymer that pinches the membrane tubule, leading to severing of the membrane tubule and formation of a membrane vesicle. Release of GDP and inorganic phosphate leads to dynamin disassembly. Following disassembly the dynamin may either dissociate from the membrane or remain associated to the vesicle and be transported to another region of the cell. Three homologous dynamin genes have been discovered, in addition to several dynamin-related proteins. Conserved dynamin regions are the N-terminal GTP-binding domain, a central pleckstrin homology domain that binds membranes, a central coiled-coil region that may activate dynamin's GTPase activity, and a C-terminal proline-rich domain that contains several motifs that bind SH3 domains on other proteins. Some dynamin-related proteins do not contain the pleckstrin homology domain or the proline-rich

domain. (See McNiven, M.A. (1998) Cell 94:151-154; Scaife, R.M. and R.L. Margolis (1997) Cell. Signal. 9:395-401.)

The cytoskeleton is reviewed in Lodish, H. et al. (1995) Molecular Cell Biology, Scientific American Books, New York NY.

Nucleic Acid-Associated Proteins

Multicellular organisms are comprised of diverse cell types that differ dramatically both in structure and function. The identity of a cell is determined by its characteristic pattern of gene expression, and different cell types express overlapping but distinctive sets of genes throughout development. Spatial and temporal regulation of gene expression is critical for the control of cell proliferation, cell differentiation, apoptosis, and other processes that contribute to organismal development. Furthermore, gene expression is regulated in response to extracellular signals that mediate cell-cell communication and coordinate the activities of different cell types. Appropriate gene regulation also ensures that cells function efficiently by expressing only those genes whose functions are required at a given time.

Transcription Factors

Transcriptional regulatory proteins are essential for the control of gene expression. Some of these proteins function as transcription factors that initiate, activate, repress, or terminate gene transcription. Transcription factors generally bind to the promoter, enhancer, and upstream regulatory regions of a gene in a sequence-specific manner, although some factors bind regulatory elements within or downstream of a gene coding region. Transcription factors may bind to a specific region of DNA singly or as a complex with other accessory factors. (Reviewed in Lewin, B. (1990) Genes IV, Oxford University Press, New York, NY, and Cell Press, Cambridge, MA, pp. 554-570.)

The double helix structure and repeated sequences of DNA create topological and chemical features which can be recognized by transcription factors. These features are hydrogen bond donor and acceptor groups, hydrophobic patches, major and minor grooves, and regular, repeated stretches of sequence which induce distinct bends in the helix. Typically, transcription factors recognize specific DNA sequence motifs of about 20 nucleotides in length. Multiple, adjacent transcription factor-binding motifs may be required for gene regulation.

Many transcription factors incorporate DNA-binding structural motifs which comprise either a helices or ß sheets that bind to the major groove of DNA. Four well-characterized structural motifs are helix-turn-helix, zinc finger, leucine zipper, and helix-loop-helix. Proteins containing these motifs may act alone as monomers, or they may form homo- or heterodimers that interact with DNA.

The helix-turn-helix motif consists of two a helices connected at a fixed angle by a short chain

of amino acids. One of the helices binds to the major groove. Helix-turn-helix motifs are exemplified by the homeobox motif which is present in homeodomain proteins. These proteins are critical for specifying the anterior-posterior body axis during development and are conserved throughout the animal kingdom. The Antennapedia and Ultrabithorax proteins of Drosophila melanogaster are prototypical homeodomain proteins. (Pabo, C.O. and R.T. Sauer (1992) Ann. Rev. Biochem. 61:1053-1095.)

The zinc finger motif, which binds zinc ions, generally contains tandem repeats of about 30 amino acids consisting of periodically spaced cysteine and histidine residues. Examples of this sequence pattern, designated C2H2 and C3HC4 ("RING" finger), have been described. (Lewin, supra.) Zinc finger proteins each contain an a helix and an antiparallel ß sheet whose proximity and conformation are maintained by the zinc ion. Contact with DNA is made by the arginine preceding the a helix and by the second, third, and sixth residues of the a helix. Variants of the zinc finger motif include poorly defined cysteine-rich motifs which bind zinc or other metal ions. These motifs may not contain histidine residues and are generally nonrepetitive. The zinc finger motif may be repeated in a tandem array within a protein, such that the a helix of each zinc finger in the protein makes contact with the major groove of the DNA double helix. This repeated contact between the protein and the DNA produces a strong and specific DNA-protein interaction. The strength and specificity of the interaction can be regulated by the number of zinc finger motifs within the protein. Though originally identified in DNA-binding proteins as regions that interact directly with DNA, zinc fingers occur in a variety of proteins that do not bind DNA (Lodish, H. et al. (1995) Molecular Cell Biology, Scientific American Books, New York, NY, pp. 447-451). For example, Galcheva-Gargova, Z. et al. (1996) Science 272:1797-1802) have identified zinc finger proteins that interact with various cytokine receptors.

The C2H2-type zinc finger signature motif contains a 28 amino acid sequence, including 2 conserved Cys and 2 conserved His residues in a C-2-C-12-H-3-H type motif. The motif generally occurs in multiple tandem repeats. A cysteine-rich domain including the motif Asp-His-His-Cys (DHHC-CRD) has been identified as a distinct subgroup of zinc finger proteins. The DHHC-CRD region has been implicated in growth and development. One DHHC-CRD mutant shows defective function of Ras, a small membrane-associated GTP-binding protein that regulates cell growth and differentiation, while other DHHC-CRD proteins probably function in pathways not involving Ras (Bartels, D.J. et al. (1999) Mol. Cell Biol. 19:6775-6787).

Zinc-finger transcription factors are often accompanied by modular sequence motifs such as the Kruppel-associated box (KRAB) and the SCAN domain. For example, the

hypoalphalipoproteinemia susceptibility gene ZNF202 encodes a SCAN box and a KRAB domain followed by eight C2H2 zinc-finger motifs (Honer, C. et al. (2001) Biochim. Biophys. Acta 1517:441-448). The SCAN domain is a highly conserved, leucine-rich motif of approximately 60 amino acids found at the amino-terminal end of zinc finger transcription factors. SCAN domains are most often linked to C2H2 zinc finger motifs through their carboxyl-terminal end. Biochemical binding studies have established the SCAN domain as a selective hetero- and homotypic oligomerization domain. SCAN domain-mediated protein complexes may function to modulate the biological function of transcription factors (Schumacher, C. et al., (2000) J. Biol. Chem. 275:17173-17179).

The KRAB (Kruppel-associated box) domain is a conserved amino acid sequence spanning approximately 75 amino acids and is found in almost one-third of the 300 to 700 genes encoding C2H2 zinc fingers. The KRAB domain is found N-terminally with respect to the finger repeats. The KRAB domain is generally encoded by two exons; the KRAB-A region or box is encoded by one exon and the KRAB-B region or box is encoded by a second exon. The function of the KRAB domain is the repression of transcription. Transcription repression is accomplished by recruitment of either the KRAB-associated protein-1, a transcriptional corepressor, or the KRAB-A interacting protein. Proteins containing the KRAB domain are likely to play a regulatory role during development (Williams, A.J. et al., (1999) Mol. Cell Biol. 19:8526-8535). A subgroup of highly related human KRAB zinc finger proteins detectable in all human tissues is highly expressed in human T lymphoid cells (Bellefroid, E.J. et al. (1993) EMBO J. 12:1363-1374). The ZNF85 KRAB zinc finger gene, a member of the human ZNF91 family, is highly expressed in normal adult testis, in seminomas, and in the NT2/D1 teratocarcinoma cell line (Poncelet, D.A. et al. (1998) DNA Cell Biol.17:931-943).

The C4 motif is found in hormone-regulated proteins. The C4 motif generally includes only 2 repeats. A number of eukaryotic and viral proteins contain a conserved cysteine-rich domain of 40 to 60 residues (called C3HC4 zinc-finger or RING finger) that binds two atoms of zinc, and is probably involved in mediating protein-protein interactions. The 3D "cross-brace" structure of the zinc ligation system is unique to the RING domain. The spacing of the cysteines in such a domain is C-x(2)-C-x(9 to 39)-C-x(1 to 3)-H-x(2 to3)-C-x(2)-C-x(4 to 48)-C-x(2)-C. The PHD finger is a C4HC3 zinc-finger-like motif found in nuclear proteins thought to be involved in chromatin-mediated transcriptional regulation.

GATA-type transcription factors contain one or two zinc finger domains which bind specifically to a region of DNA that contains the consecutive nucleotide sequence GATA. NMR studies indicate that the zinc finger comprises two irregular anti-parallel b sheets and an a helix, followed by a long loop to the C-terminal end of the finger (Ominchinski, J.G. (1993) Science

261:438-446). The helix and the loop connecting the two b-sheets contact the major groove of the DNA, while the C-terminal part, which determines the specificity of binding, wraps around into the minor groove.

The LIM motif consists of about 60 amino acid residues and contains seven conserved cysteine residues and a histidine within a consensus sequence (Schmeichel, K.L. and Beckerle, M.C. (1994) Cell 79:211-219). The LIM family includes transcription factors and cytoskeletal proteins which may be involved in development, differentiation, and cell growth. One example is actin-binding LIM protein, which may play roles in regulation of the cytoskeleton and cellular morphogenesis (Roof, D.J. et al. (1997) J. Cell Biol. 138:575-588). The N-terminal domain of actin-binding LIM protein has four double zinc finger motifs with the LIM consensus sequence. The C-terminal domain of actin-binding LIM protein shows sequence similarity to known actin-binding proteins such as dematin and villin. Actin-binding LIM protein binds to F-actin through its dematin-like C-terminal domain. The LIM domain may mediate protein-protein interactions with other LIM-binding proteins.

Myeloid cell development is controlled by tissue-specific transcription factors. Myeloid zinc finger proteins (MZF) include MZF-1 and MZF-2. MZF-1 functions in regulation of the development of neutrophilic granulocytes. A murine homolog MZF-2 is expressed in myeloid cells, particularly in the cells committed to the neutrophilic lineage. MZF-2 is down-regulated by G-CSF and appears to have a unique function in neutrophil development (Murai, K. et al. (1997) Genes Cells 2:581-591).

The leucine zipper motif comprises a stretch of amino acids rich in leucine which can form an amphipathic a helix. This structure provides the basis for dimerization of two leucine zipper proteins. The region adjacent to the leucine zipper is usually basic, and upon protein dimerization, is optimally positioned for binding to the major groove. Proteins containing such motifs are generally referred to as bZIP transcription factors. The leucine zipper motif is found in the proto-oncogenes Fos and Jun, which comprise the heterodimeric transcription factor AP1 involved in cell growth and the determination of cell lineage (Papavassiliou, A. G. (1995) N. Engl. J. Med. 332:45-47).

The helix-loop-helix motif (HLH) consists of a short a helix connected by a loop to a longer a helix. The loop is flexible and allows the two helices to fold back against each other and to bind to DNA. The transcription factor Myc contains a prototypical HLH motif.

The NF-kappa-B/Rel signature defines a family of eukaryotic transcription factors involved in oncogenesis, embryonic development, differentiation and immune response. Most transcription factors containing the Rel homology domain (RHD) bind as dimers to a consensus DNA sequence motif termed kappa-B. Members of the Rel family share a highly conserved 300 amino acid domain termed the Rel homology domain. The characteristic Rel C-terminal domain is involved in gene activation and

cytoplasmic anchoring functions. Proteins known to contain the RHD domain include vertebrate nuclear factor NF-kappa-B, which is a heterodimer of a DNA-binding subunit and the transcription factor p65, mammalian transcription factor RelB, and vertebrate proto-oncogene c-rel, a protein associated with differentiation and lymphopoiesis (Kabrun, N., and Enrietto, P.J. (1994) Semin. Cancer Biol. 5:103-112).

A DNA binding motif termed ARID (AT-rich interactive domain) distinguishes an evolutionarily conserved family of proteins. The approximately 100-residue ARID sequence is present in a series of proteins strongly implicated in the regulation of cell growth, development, and tissue-specific gene expression. ARID proteins include Bright (a regulator of B-cell-specific gene expression), dead ringer (involved in development), and MRF-2 (which represses expression from the cytomegalovirus enhancer) (Dallas, P.B. et al. (2000) Mol. Cell Biol. 20:3137-3146).

The ELM2 (Egl-27 and MTA1 homology 2) domain is found in metastasis-associated protein MTA1 and protein ER1. The Caenorhabditis elegans gene egl-27 is required for embryonic patterning MTA1, a human gene with elevated expression in metastatic carcinomas, is a component of a protein complex with histone deacetylase and nucleosome remodelling activities (Solari, F. et al. (1999) Development 126:2483-2494). The ELM2 domain is usually found to the N terminus of a myb-like DNA binding domain. ELM2 is also found associated with an ARID DNA.

Most transcription factors contain characteristic DNA binding motifs, and variations on the above motifs and new motifs have been and are currently being characterized. (Faisst, S. and S. Meyer (1992) Nucl. Acids Res. 20:3-26.)

Chromatin Associated Proteins

In the nucleus, DNA is packaged into chromatin, the compact organization of which limits the accessibility of DNA to transcription factors and plays a key role in gene regulation. (Lewin, supra, pp. 409-410.) The compact structure of chromatin is determined and influenced by chromatin-associated proteins such as the histones, the high mobility group (HMG) proteins, and the chromodomain proteins. There are five classes of histones, H1, H2A, H2B, H3, and H4, all of which are highly basic, low molecular weight proteins. The fundamental unit of chromatin, the nucleosome, consists of 200 base pairs of DNA associated with two copies each of H2A, H2B, H3, and H4. H1 links adjacent nucleosomes. HMG proteins are low molecular weight, non-histone proteins that may play a role in unwinding DNA and stabilizing single-stranded DNA. Chromodomain proteins play a key role in the formation of highly compacted heterochromatin, which is transcriptionally silent.

Diseases and Disorders Related to Gene Regulation

Many neoplastic disorders in humans can be attributed to inappropriate gene expression.

Malignant cell growth may result from either excessive expression of tumor promoting genes or insufficient expression of tumor suppressor genes. (Cleary, M.L. (1992) Cancer Surv. 15:89-104.) The zinc finger-type transcriptional regulator WT1 is a tumor-suppressor protein that is inactivated in children with Wilm's tumor. The oncogene bcl-6, which plays an important role in large-cell lymphoma, is also a zinc-finger protein (Papavassiliou, A. G. (1995) N. Engl. J. Med. 332:45-47). Chromosomal translocations may also produce chimeric loci that fuse the coding sequence of one gene with the regulatory regions of a second unrelated gene. Such an arrangement likely results in inappropriate gene transcription, potentially contributing to malignancy. In Burkitt's lymphoma, for example, the transcription factor Myc is translocated to the immunoglobulin heavy chain locus, greatly enhancing Myc expression and resulting in rapid cell growth leading to leukemia (Latchman, D. S. (1996) N. Engl. J. Med. 334:28-33).

In addition, the immune system responds to infection or trauma by activating a cascade of events that coordinate the progressive selection, amplification, and mobilization of cellular defense mechanisms. A complex and balanced program of gene activation and repression is involved in this process. However, hyperactivity of the immune system as a result of improper or insufficient regulation of gene expression may result in considerable tissue or organ damage. This damage is well-documented in immunological responses associated with arthritis, allergens, heart attack, stroke, and infections. (Isselbacher et al. Harrison's Principles of Internal Medicine, 13/e, McGraw Hill, Inc. and Teton Data Systems Software, 1996.) The causative gene for autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED) was recently isolated and found to encode a protein with two PHD-type zinc finger motifs (Bjorses, P. et al. (1998) Hum. Mol. Genet. 7:1547-1553).

Furthermore, the generation of multicellular organisms is based upon the induction and coordination of cell differentiation at the appropriate stages of development. Central to this process is differential gene expression, which confers the distinct identities of cells and tissues throughout the body. Failure to regulate gene expression during development could result in developmental disorders. Human developmental disorders caused by mutations in zinc finger-type transcriptional regulators include: urogenital developmental abnormalities associated with WT1; Greig cephalopolysyndactyly, Pallister-Hall syndrome, and postaxial polydactyly type A (GLI3), and Townes-Brocks syndrome, characterized by anal, renal, limb, and ear abnormalities (SALL1) (Engelkamp, D. and V. van Heyningen (1996) Curr. Opin. Genet. Dev. 6:334-342; Kohlhase, J. et al. (1999) Am. J. Hum. Genet. 64:435-445).

SYNTHESIS OF NUCLEIC ACIDS

Polymerases

DNA and RNA replication are critical processes for cell replication and function. DNA and RNA replication are mediated by the enzymes DNA and RNA polymerase, respectively, by a "templating" process in which the nucleotide sequence of a DNA or RNA strand is copied by complementary base-pairing into a complementary nucleic acid sequence of either DNA or RNA. However, there are fundamental differences between the two processes.

DNA polymerase catalyzes the stepwise addition of a deoxyribonucleotide to the 3'-OH end of a polynucleotide strand (the primer strand) that is paired to a second (template) strand. The new DNA strand therefore grows in the 5' to 3' direction (Alberts, B. et al. (1994) The Molecular Biology of the Cell, Garland Publishing Inc., New York, NY, pp 251-254). The substrates for the polymerization reaction are the corresponding deoxynucleotide triphosphates which must base-pair with the correct nucleotide on the template strand in order to be recognized by the polymerase. Because DNA exists as a double-stranded helix, each of the two strands may serve as a template for the formation of a new complementary strand. Each of the two daughter cells of a dividing cell therefore inherits a new DNA double helix containing one old and one new strand. Thus, DNA is said to be replicated "semiconservatively" by DNA polymerase. In addition to the synthesis of new DNA, DNA polymerase is also involved in the repair of damaged DNA as discussed below under "Ligases."

In contrast to DNA polymerase, RNA polymerase uses a DNA template strand to "transcribe" DNA into RNA using ribonucleotide triphosphates as substrates. Like DNA polymerization, RNA polymerization proceeds in a 5' to 3' direction by addition of a ribonucleoside monophosphate to the 3'-OH end of a growing RNA chain. DNA transcription generates messenger RNAs (mRNA) that carry information for protein synthesis, as well as the transfer, ribosomal, and other RNAs that have structural or catalytic functions. In eukaryotes, three discrete RNA polymerases synthesize the three different types of RNA (Alberts et al., supra pp. 367-368). RNA polymerase I makes the large ribosomal RNAs, RNA polymerase II makes the mRNAs that will be translated into proteins, and RNA polymerase III makes a variety of small, stable RNAs, including 5S ribosomal RNA and the transfer RNAs (tRNA). In all cases, RNA synthesis is initiated by binding of the RNA polymerase to a promoter region on the DNA and synthesis begins at a start site within the promoter. Synthesis is completed at a stop (termination) signal in the DNA whereupon both the polymerase and the completed RNA chain are released.

Ligases

DNA repair is the process by which accidental base changes, such as those produced by oxidative damage, hydrolytic attack, or uncontrolled methylation of DNA, are corrected before replication or transcription of the DNA can occur. Because of the efficiency of the DNA repair

process, fewer than one in a thousand accidental base changes causes a mutation (Alberts et al., supra pp. 245-249). The three steps common to most types of DNA repair are (1) excision of the damaged or altered base or nucleotide by DNA nucleases, (2) insertion of the correct nucleotide in the gap left by the excised nucleotide by DNA polymerase using the complementary strand as the template and, (3) sealing the break left between the inserted nucleotide(s) and the existing DNA strand by DNA ligase. In the last reaction, DNA ligase uses the energy from ATP hydrolysis to activate the 5' end of the broken phosphodiester bond before forming the new bond with the 3'-OH of the DNA strand. In Bloom's syndrome, an inherited human disease, individuals are partially deficient in DNA ligation and consequently have an increased incidence of cancer (Alberts et al., supra p. 247).

Nucleases

Nucleases comprise enzymes that hydrolyze both DNA (DNase) and RNA (Rnase). They serve different purposes in nucleic acid metabolism. Nucleases hydrolyze the phosphodiester bonds between adjacent nucleotides either at internal positions (endonucleases) or at the terminal 3' or 5' nucleotide positions (exonucleases). A DNA exonuclease activity in DNA polymerase, for example, serves to remove improperly paired nucleotides attached to the 3'-OH end of the growing DNA strand by the polymerase and thereby serves a "proofreading" function. As mentioned above, DNA endonuclease activity is involved in the excision step of the DNA repair process.

RNases also serve a variety of functions. For example, RNase P is a ribonucleoprotein enzyme which cleaves the 5' end of pre-tRNAs as part of their maturation process. RNase H digests the RNA strand of an RNA/DNA hybrid. Such hybrids occur in cells invaded by retroviruses, and RNase H is an important enzyme in the retroviral replication cycle. Pancreatic RNase secreted by the pancreas into the intestine hydrolyzes RNA present in ingested foods. RNase activity in serum and cell extracts is elevated in a variety of cancers and infectious diseases (Schein, C.H. (1997) Nat. Biotechnol. 15:529-536). Regulation of RNase activity is being investigated as a means to control tumor angiogenesis, allergic reactions, viral infection and replication, and fungal infections.

MODIFICATION OF NUCLEIC ACIDS

Methylases

Methylation of specific nucleotides occurs in both DNA and RNA, and serves different functions in the two macromolecules. Methylation of cytosine residues to form 5-methyl cytosine in DNA occurs specifically in CG sequences which are base-paired with one another in the DNA double-helix. The pattern of methylation is passed from generation to generation during DNA replication by an enzyme called "maintenance methylase" that acts preferentially on those CG sequences that are base-paired with a CG sequence that is already methylated. Such methylation

appears to distinguish active from inactive genes by preventing the binding of regulatory proteins that "turn on" the gene, but permiting the binding of proteins that inactivate the gene (Alberts et al. supra pp. 448-451). In RNA metabolism, "tRNA methylase" produces one of several nucleotide modifications in tRNA that affect the conformation and base-pairing of the molecule and facilitate the recognition of the appropriate mRNA codons by specific tRNAs. The primary methylation pattern is the dimethylation of guanine residues to form N,N-dimethyl guanine.

Helicases and Single-stranded Binding Proteins

Helicases are enzymes that destabilize and unwind double helix structures in both DNA and RNA. Since DNA replication occurs more or less simultaneously on both strands, the two strands must first separate to generate a replication "fork" for DNA polymerase to act on. Two types of replication proteins contribute to this process, DNA helicases and single-stranded binding proteins. DNA helicases hydrolyze ATP and use the energy of hydrolysis to separate the DNA strands. Single-stranded binding proteins (SSBs) then bind to the exposed DNA strands, without covering the bases, thereby temporarily stabilizing them for templating by the DNA polymerase (Alberts et al. supra pp. 255-256).

RNA helicases also alter and regulate RNA conformation and secondary structure. Like the DNA helicases, RNA helicases utilize energy derived from ATP hydrolysis to destabilize and unwind RNA duplexes. The most well-characterized and ubiquitous family of RNA helicases is the DEAD-box family, so named for the conserved B-type ATP-binding motif which is diagnostic of proteins in this family. Over 40 DEAD-box helicases have been identified in organisms as diverse as bacteria, insects, yeast, amphibians, mammals, and plants. DEAD-box helicases function in diverse processes such as translation initiation, splicing, ribosome assembly, and RNA editing, transport, and stability. Examples of these RNA helicases include yeast Drs1 protein, which is involved in ribosomal RNA processing; yeast TIF1 and TIF2 and mammalian eIF-4A, which are essential to the initiation of RNA translation; and human p68 antigen, which regulates cell growth and division (Ripmaster, T.L. et al. (1992) Proc. Natl. Acad. Sci. USA 89:11131-11135; Chang, T.-H. et al. (1990) Proc. Natl. Acad. Sci. USA 87:1571-1575). These RNA helicases demonstrate strong sequence homology over a stretch of some 420 amino acids. Included among these conserved sequences are the consensus sequence for the A motif of an ATP binding protein; the "DEAD box" sequence, associated with ATPase activity; the sequence SAT, associated with the actual helicase unwinding region; and an octapeptide consensus sequence, required for RNA binding and ATP hydrolysis (Pause, A. et al. (1993) Mol. Cell Biol. 13:6789-6798). Differences outside of these conserved regions are believed to reflect differences in the functional roles of individual proteins (Chang, T.H. et al. (1990) Proc. Natl. Acad.

Sci. USA 87:1571-1575).

Some DEAD-box helicases play tissue- and stage-specific roles in spermatogenesis and embryogenesis. Overexpression of the DEAD-box 1 protein (DDX1) may play a role in the progression of neuroblastoma (Nb) and retinoblastoma (Rb) tumors (Godbout, R. et al. (1998) J. Biol. Chem. 273:21161-21168). These observations suggest that DDX1 may promote or enhance tumor progression by altering the normal secondary structure and expression levels of RNA in cancer cells. Other DEAD-box helicases have been implicated either directly or indirectly in tumorigenesis. (Discussed in Godbout, supra.) For example, murine p68 is mutated in ultraviolet light-induced tumors, and human DDX6 is located at a chromosomal breakpoint associated with B-cell lymphoma. Similarly, a chimeric protein comprised of DDX10 and NUP98, a nucleoporin protein, may be involved in the pathogenesis of certain myeloid malignancies.

Topoisomerases

Besides the need to separate DNA strands prior to replication, the two strands must be "unwound" from one another prior to their separation by DNA helicases. This function is performed by proteins known as DNA topoisomerases. DNA topoisomerase effectively acts as a reversible nuclease that hydrolyzes a phosphodiesterase bond in a DNA strand, permits the two strands to rotate freely about one another to remove the strain of the helix, and then rejoins the original phosphodiester bond between the two strands. Topoisomerases are essential enzymes responsible for the topological rearrangement of DNA brought about by transcription, replication, chromatin formation, recombination, and chromosome segregation. Superhelical coils are introduced into DNA by the passage of processive enzymes such as RNA polymerase, or by the separation of DNA strands by a helicase prior to replication. Knotting and concatenation can occur in the process of DNA synthesis, storage, and repair. All topoisomerases work by breaking a phosphodiester bond in the ribose-phosphate backbone of DNA. A catalytic tyrosine residue on the enzyme makes a nucleophilic attack on the scissile phosphodiester bond, resulting in a reaction intermediate in which a covalent bond is formed between the enzyme and one end of the broken strand. A tyrosine-DNA phosphodiesterase functions in DNA repair by hydrolyzing this bond in occasional dead-end topoisomerase I-DNA intermediates (Pouliot, J.J. et al. (1999) Science 286:552-555).

Two types of DNA topoisomerase exist, types I and II. Type I topoisomerases work as monomers, making a break in a single strand of DNA while type II topoisomerases, working as homodimers, cleave both strands. DNA Topoisomerase I causes a single-strand break in a DNA helix to allow the rotation of the two strands of the helix about the remaining phosphodiester bond in the opposite strand. DNA topoisomerase II causes a transient break in both strands of a DNA helix

where two double helices cross over one another. This type of topoisomerase can efficiently separate two interlocked DNA circles (Alberts et al. supra pp.260-262). Type II topoisomerases are largely confined to proliferating cells in eukaryotes, such as cancer cells. For this reason they are targets for anticancer drugs. Topoisomerase II has been implicated in multi-drug resistance (MDR) as it appears to aid in the repair of DNA damage inflicted by DNA binding agents such as doxorubicin and vincristine.

The topoisomerase I family includes topoisomerases I and III (topo I and topo III). The crystal structure of human topoisomerase I suggests that rotation about the intact DNA strand is partially controlled by the enzyme. In this "controlled rotation" model, protein-DNA interactions limit the rotation, which is driven by torsional strain in the DNA (Stewart, L. et al. (1998) Science 379:1534-1541). Structurally, topo I can be recognized by its catalytic tyrosine residue and a number of other conserved residues in the active site region. Topo I is thought to function during transcription. Two topo IIIs are known in humans, and they are homologous to prokaryotic topoisomerase I, with a conserved tyrosine and active site signature specific to this family. Topo III has been suggested to play a role in meiotic recombination. A mouse topo III is highly expressed in testis tissue and its expression increases with the increase in the number of cells in pachytene (Seki, T. et al. (1998) J. Biol. Chem. 273:28553-28556).

The topoisomerase II family includes two isozymes (IIa and IIb) encoded by different genes. Topo II cleaves double stranded DNA in a reproducible, nonrandom fashion, preferentially in an AT rich region, but the basis of cleavage site selectivity is not known. Structurally, topo II is made up of four domains, the first two of which are structurally similar and probably distantly homologous to similar domains in eukaryotic topo I. The second domain bears the catalytic tyrosine, as well as a highly conserved pentapeptide. The IIa isoform appears to be responsible for unlinking DNA during chromosome segregation. Cell lines expressing IIa but not IIb suggest that IIb is dispensable in cellular processes; however, IIb knockout mice died perinatally due to a failure in neural development. That the major abnormalities occurred in predominantly late developmental events (neurogenesis) suggests that IIb is needed not at mitosis, but rather during DNA repair (Yang, X. et al. (2000) Science 287:131-134).

Topoisomerases have been implicated in a number of disease states, and topoisomerase poisons have proven to be effective anti-tumor drugs for some human malignancies. Topo I is mislocalized in Fanconi's anemia, and may be involved in the chromosomal breakage seen in this disorder (Wunder, E. (1984) Hum. Genet. 68:276-281). Overexpression of a truncated topo III in ataxia-telangiectasia (A-T) cells partially suppresses the A-T phenotype, probably through a dominant

negative mechanism. This suggests that topo III is deregulated in A-T (Fritz, E. et al. (1997) Proc. Natl. Acad. Sci. USA 94:4538-4542). Topo III also interacts with the Bloom's Syndrome gene product, and has been suggested to have a role as a tumor suppressor (Wu, L. et al. (2000) J. Biol. Chem. 275:9636-9644). Aberrant topo II activity is often associated with cancer or increased cancer risk. Greatly lowered topo II activity has been found in some, but not all A-T cell lines (Mohamed, R. et al. (1987) Biochem. Biophys. Res. Commun. 149:233-238). On the other hand, topo II can break DNA in the region of the A-T gene (ATM), which controls all DNA damage-responsive cell cycle checkpoints (Kaufmann, W.K. (1998) Proc. Soc. Exp. Biol. Med. 217:327-334). The ability of topoisomerases to break DNA has been used as the basis of antitumor drugs. Topoisomerase poisons act by increasing the number of dead-end covalent DNA-enzyme complexes in the cell, ultimately triggering cell death pathways (Fortune, J.M. and N. Osheroff (2000) Prog. Nucleic Acid Res. Mol. Biol. 64:221-253; Guichard, S.M. and M.K. Danks (1999) Curr. Opin. Oncol. 11:482-489). Antibodies against topo I are found in the serum of systemic sclerosis patients, and the levels of the antibody may be used as a marker of pulmonary involvement in the disease (Diot, E. et al. (1999) Chest 116:715-720). Finally, the DNA binding region of human topo I has been used as a DNA delivery vehicle for gene therapy (Chen, T.Y. et al. (2000) Appl. Microbiol. Biotechnol. 53:558-567).

Recombinases

Genetic recombination is the process of rearranging DNA sequences within an organism's genome to provide genetic variation for the organism in response to changes in the environment. DNA recombination allows variation in the particular combination of genes present in an individual's genome, as well as the timing and level of expression of these genes. (See Alberts et al. supra pp. 263-273.) Two broad classes of genetic recombination are commonly recognized, general recombination and site-specific recombination. General recombination involves genetic exchange between any homologous pair of DNA sequences usually located on two copies of the same chromosome. The process is aided by enzymes, recombinases, that "nick" one strand of a DNA duplex more or less randomly and permit exchange with a complementary strand on another duplex. The process does not normally change the arrangement of genes in a chromosome. In site-specific recombination, the recombinase recognizes specific nucleotide sequences present in one or both of the recombining molecules. Base-pairing is not involved in this form of recombination and therefore it does not require DNA homology between the recombining molecules. Unlike general recombination, this form of recombination can alter the relative positions of nucleotide sequences in chromosomes.

RNA METABOLISM

Ribonucleic acid (RNA) is a linear single-stranded polymer of four nucleotides, ATP, CTP,

UTP, and GTP. In most organisms, RNA is transcribed as a copy of deoxyribonucleic acid (DNA), the genetic material of the organism. In retroviruses RNA rather than DNA serves as the genetic material. RNA copies of the genetic material encode proteins or serve various structural, catalytic, or regulatory roles in organisms. RNA is classified according to its cellular localization and function. Messenger RNAs (mRNAs) encode polypeptides. Ribosomal RNAs (rRNAs) are assembled, along with ribosomal proteins, into ribosomes, which are cytoplasmic particles that translate mRNA into polypeptides. Transfer RNAs (tRNAs) are cytosolic adaptor molecules that function in mRNA translation by recognizing both an mRNA codon and the amino acid that matches that codon. Heterogeneous nuclear RNAs (hnRNAs) include mRNA precursors and other nuclear RNAs of various sizes. Small nuclear RNAs (snRNAs) are a part of the nuclear spliceosome complex that removes intervening, non-coding sequences (introns) and rejoins exons in pre-mRNAs.

Proteins are associated with RNA during its transcription from DNA, RNA processing, and translation of mRNA into protein. Proteins are also associated with RNA as it is used for structural, catalytic, and regulatory purposes.

RNA Processing

Ribosomal RNAs (rRNAs) are assembled, along with ribosomal proteins, into ribosomes, which are cytoplasmic particles that translate messenger RNA (mRNA) into polypeptides. The eukaryotic ribosome is composed of a 60S (large) subunit and a 40S (small) subunit, which together form the 80S ribosome. In addition to the 18S, 28S, 5S, and 5.8S rRNAs, ribosomes contain from 50 to over 80 different ribosomal proteins, depending on the organism. Ribosomal proteins are classified according to which subunit they belong (i.e., L, if associated with the large 60S large subunit or S if associated with the small 40S subunit). E. coli ribosomes have been the most thoroughly studied and contain 50 proteins, many of which are conserved in all life forms. The structures of nine ribosomal proteins have been solved to less than 3.0D resolution (i.e., S5, S6, S17, L1, L6, L9, L12, L14, L30), revealing common motifs, such as b-a- b protein folds in addition to acidic and basic RNA-binding motifs positioned between b-strands. Most ribosomal proteins are believed to contact rRNA directly (reviewed in Liljas, A. and Garber, M. (1995) Curr. Opin. Struct. Biol. 5:721–727; see also Woodson, S.A. and Leontis, N.B. (1998) Curr. Opin. Struct. Biol. 8:294-300; Ramakrishnan, V. and White, S.W. (1998) Trends Biochem. Sci. 23:208-212).

Ribosomal proteins may undergo post-translational modifications or interact with other ribosome-associated proteins to regulate translation. For example, the highly homologous 40S ribosomal protein S6 kinases (S6K1 and S6K2) play a key role in the regulation of cell growth by controlling the biosynthesis of translational components which make up the protein synthetic apparatus

(including the ribosomal proteins). In the case of S6K1, at least eight phosphorylation sites are believed to mediate kinase activation in a hierarchical fashion (Dufner and Thomas (1999) Exp. Cell. Res. 253:100-109). Some of the ribosomal proteins, including L1, also function as translational repressors by binding to polycistronic mRNAs encoding ribosomal proteins (reviewed in Liljas, A. supra and Garber, M. supra).

Recent evidence suggests that a number of ribosomal proteins have secondary functions independent of their involvement in protein biosynthesis. These proteins function as regulators of cell proliferation and, in some instances, as inducers of cell death. For example, the expression of human ribosomal protein L13a has been shown to induce apoptosis by arresting cell growth in the G2/M phase of the cell cycle. Inhibition of expression of L13a induces apoptosis in target cells, which suggests that this protein is necessary, in the appropriate amount, for cell survival. Similar results have been obtained in yeast where inactivation of yeast homologues of L13a, rp22 and rp23, results in severe growth retardation and death. A closely related ribosomal protein, L7, arrests cells in G1 and also induces apoptosis. Thus, it appears that a subset of ribosomal proteins may function as cell cycle checkpoints and compose a new family of cell proliferation regulators.

Mapping of individual ribosomal proteins on the surface of intact ribosomes is accomplished using 3D immunocryoelectronmicroscopy, whereby antibodies raised against specific ribosomal proteins are visualized. Progress has been made toward the mapping of L1, L7, and L12 while the structure of the intact ribosome has been solved to only 20-25D resolution and inconsistencies exist among different crude structures (Frank, J. (1997) Curr. Opin. Struct. Biol. 7:266–272).

Three distinct sites have been identified on the ribosome. The aminoacyl-tRNA acceptor site (A site) receives charged tRNAs (with the exception of the initiator-tRNA). The peptidyl-tRNA site (P site) binds the nascent polypeptide as the amino acid from the A site is added to the elongating chain. Deacylated tRNAs bind in the exit site (E site) prior to their release from the ribosome. The structure of the ribosome is reviewed in Stryer, L. (1995) Biochemistry W.H. Freeman and Company, New York NY pp. 888-9081; Lodish, H. et al. (1995) Molecular Cell Biology Scientific American Books, New York NY pp. 119-138; and Lewin, B (1997) Genes VI Oxford University Press, Inc. New York, NY).

Various proteins are necessary for processing of transcribed RNAs in the nucleus. Pre-mRNA processing steps include capping at the 5' end with methylguanosine, polyadenylating the 3' end, and splicing to remove introns. The primary RNA transript from DNA is a faithful copy of the gene containing both exon and intron sequences, and the latter sequences must be cut out of the RNA transcript to produce a mRNA that codes for a protein. This "splicing" of the mRNA sequence takes

place in the nucleus with the aid of a large, multicomponent ribonucleoprotein complex known as a spliceosome. The spliceosomal complex is comprised of five small nuclear ribonucleoprotein particles (snRNPs) designated U1, U2, U4, U5, and U6. Each snRNP contains a single species of snRNA and about ten proteins. The RNA components of some snRNPs recognize and base-pair with intron consensus sequences. The protein components mediate spliceosome assembly and the splicing reaction. Autoantibodies to snRNP proteins are found in the blood of patients with systemic lupus erythematosus (Stryer, L. (1995) Biochemistry W.H. Freeman and Company, New York NY, p. 863).

Several splicing regulatory proteins have been identified in Drosophila. Human (HsSWAP) and mouse (MmSWAP) homologs of the suppressor-of-white-apricot (su(wa)) gene have been cloned and characterized. HsSWAP and MmSWAP both have five highly homologous regions to su(wa), including an arginine/serine-rich domain and two repeated modules that are homologous to regions in the constitutive splicing factor, SPP91/PRP21. Mammalian SWAP mRNAs are alternatively spliced at the same splice sites as in Drosophila. The splice junctions of the Drosophila and mammalian regulated introns are conserved. Thus, research suggests that the mammalian SWAP gene functions as a vertebrate alternative splicing regulator (Denhez, F. and Lafyatis, R. (1994) Biol. Chem. 269:16170-16179).

Serine- and arginine-rich pre-mRNA splicing factors (SR proteins) are phosphorylated before they regulate splicing events. SRrp86 (SR-related protein of 86 kDa) is a novel SR protein containing a single amino-terminal RNA recognition motif and two carboxy-terminal domains rich in serine-arginine (SR) dipeptides. SRrp86 activates splicing in the presence of SRp20. However, it inhibits the in vitro and in vivo activation of specific splice sites by SR proteins, including ASF/SF2, SC35, and SRp55. Research suggests that pairwise combination of SRrp86 with specific SR proteins leads to altered splicing efficiency and differential splice site selection (Barnard, D.C. and Patton, J.G. (2000) Mol. Cell. Biol. 20:3049-3057).

Heterogeneous nuclear ribonucleoproteins (hnRNPs) have been identified that have roles in splicing, exporting of the mature RNAs to the cytoplasm, and mRNA translation (Biamonti, G. et al. (1998) Clin. Exp. Rheumatol. 16:317-326). Some examples of hnRNPs include the yeast proteins Hrp1p, involved in cleavage and polyadenylation at the 3' end of the RNA; Cbp80p, involved in capping the 5' end of the RNA; and Npl3p, a homolog of mammalian hnRNP A1, involved in export of mRNA from the nucleus (Shen, E.C. et al. (1998) Genes Dev. 12:679-691). HnRNPs have been shown to be
important targets of the autoimmune response in rheumatic diseases (Biamonti, supra).

Many snRNP and hnRNP proteins are characterized by an RNA recognition motif (RRM).

(Reviewed in Birney, E. et al. (1993) Nucleic Acids Res. 21:5803-5816.) The RRM is about 80 amino acids in length and forms four b-strands and two a-helices arranged in an a /b sandwich. The RRM contains a core RNP-1 octapeptide motif along with surrounding conserved sequences. In addition to snRNP proteins, examples of RNA-binding proteins which contain the above motifs include heteronuclear ribonucleoproteins which stabilize nascent RNA and factors which regulate alternative splicing. Alternative splicing factors include developmentally regulated proteins, specific examples of which have been identified in lower eukaryotes such as Drosophila melanogaster and Caenorhabditis elegans. These proteins play key roles in developmental processes such as pattern formation and sex determination, respectively. (See, for example, Hodgkin, J. et al. (1994) Development 120:3681-3689.)

The 3' ends of most eukaryote mRNAs are also posttranscriptionally modified by polyadenylation. Polyadenylation proceeds through two enzymatically distinct steps: (i) the endonucleolytic cleavage of nascent mRNAs at *cis*-acting polyadenylation signals in the 3'-untranslated (non-coding) region and (ii) the addition of a poly(A) tract to the 5' mRNA fragment. The presence of *cis*-acting RNA sequences is necessary for both steps. These sequences include 5'-AAUAAA-3' located 10-30 nucleotides upstream of the cleavage site and a less well-conserved GU- or U-rich sequence element located 10-30 nucleotides downstream of the cleavage site. Cleavage stimulation factor (CstF), cleavage factor I (CF I), and cleavage factor II (CF II) are involved in the cleavage reaction while cleavage and polyadenylation specificity factor (CPSF) and poly(A) polymerase (PAP) are necessary for both cleavage and polyadenylation. An additional enzyme, poly(A)-binding protein II (PAB II), promotes poly(A) tract elongation (Rüegsegger, U. et al. (1996) J. Biol. Chem. 271:6107-6113; and references within).

TRANSLATION

Correct translation of the genetic code depends upon each amino acid forming a linkage with the appropriate transfer RNA (tRNA). The aminoacyl-tRNA synthetases (aaRSs) are essential proteins found in all living organisms. The aaRSs are responsible for the activation and correct attachment of an amino acid with its cognate tRNA, as the first step in protein biosynthesis. Prokaryotic organisms have at least twenty different types of aaRSs, one for each different amino acid, while eukaryotes usually have two aaRSs, a cytosolic form and a mitochondrial form, for each different amino acid. The 20 aaRS enzymes can be divided into two structural classes. Class I enzymes add amino acids to the 2' hydroxyl at the 3' end of tRNAs while Class II enzymes add amino acids to the 3' hydroxyl at the 3' end of tRNAs. Each class is characterized by a distinctive topology of the catalytic domain. Class I enzymes contain a catalytic domain based on the nucleotide-binding

Rossman 'fold'. In particular, a consensus tetrapeptide motif is highly conserved (Prosite Document PDOC00161, Aminoacyl-transfer RNA synthetases class-I signature). Class I enzymes are specific for arginine, cysteine, glutamic acid, glutamine, isoleucine, leucine, methionine, tyrosine, tryptophan, and valine. Class II enzymes contain a central catalytic domain, which consists of a seven-stranded antiparallel ß-sheet domain, as well as N- and C- terminal regulatory domains. Class II enzymes are separated into two groups based on the heterodimeric or homodimeric structure of the enzyme; the latter group is further subdivided by the structure of the N- and C-terminal regulatory domains (Hartlein, M. and Cusack, S. (1995) J. Mol. Evol. 40:519-530). Class II enzymes are specific for alanine, asparagine, aspartic acid, glycine, histidine, lysine, phenylalanine, proline, serine, and threonine.

Certain aaRSs also have editing functions. IleRS, for example, can misactivate valine to form Val-$tRNA^{Ile}$, but this product is cleared by a hydrolytic activity that destroys the mischarged product. This editing activity is located within a second catalytic site found in the connective polypeptide 1 region (CP1), a long insertion sequence within the Rossman fold domain of Class I enzymes (Schimmel, P. et al. (1998) FASEB J. 12:1599-1609). AaRSs also play a role in tRNA processing. It has been shown that mature tRNAs are charged with their respective amino acids in the nucleus before export to the cytoplasm, and charging may serve as a quality control mechanism to insure the tRNAs are functional (Martinis, S.A. et al. (1999) EMBO J. 18:4591-4596).

Under optimal conditions, polypeptide synthesis proceeds at a rate of approximately 40 amino acid residues per second. The rate of misincorporation during translation in on the order of 10^{-4} and is primarily the result of aminoacyl-t-RNAs being charged with the incorrect amino acid. Incorrectly charged tRNA are toxic to cells as they result in the incorporation of incorrect amino acid residues into an elongating polypeptide. The rate of translation is presumed to be a compromise between the optimal rate of elongation and the need for translational fidelity. Mathematical calculations predict that 10^{-4} is indeed the maximum acceptable error rate for protein synthesis in a biological system (reviewed in Stryer, L. supra and Watson, J. et al. (1987) The Benjamin/Cummings Publishing Co., Inc. Menlo Park, CA). A particularly error prone aminoacyl-tRNA charging event is the charging of $tRNA^{Gln}$ with Gln. A mechanism exits for the correction of this mischarging event which likely has its origins in evolution. Gln was among the last of the 20 naturally occurring amino acids used in polypeptide synthesis to appear in nature. Gram positive eubacteria, cyanobacteria, Archeae, and eukaryotic organelles possess a noncanonical pathway for the synthesis of Gln-$tRNA^{Gln}$ based on the transformation of Glu-$tRNA^{Gln}$ (synthesized by Glu-tRNA synthetase, GluRS) using the enzyme Glu-$tRNA^{Gln}$ amidotransferase (Glu-AdT). The reactions involved in the transamidation pathway are as follows (Curnow, A.W. et al. (1997) Nucleic Acids Symposium 36:2-4):

$$\text{tRNA}^{\text{Gln}} + \text{Glu} + \text{ATP} \xrightarrow{\text{GluRS}} \text{Glu-tRNA}^{\text{Gln}} + \text{AMP} + \text{PP}_i$$

$$\text{Glu-tRNA}^{\text{Gln}} + \text{Gln} + \text{ATP} \xrightarrow{\text{Glu-AdT}} \text{Gln-tRNA}^{\text{Gln}} + \text{Glu} + \text{ADP} + \text{P}$$

A similar enzyme, Asp-tRNA$^{\text{Asn}}$ amidotransferase, exists in Archaea, which transforms Asp-tRNA$^{\text{Asn}}$ to Asn-tRNA$^{\text{Asn}}$. Formylase, the enzyme that transforms Met-tRNA$^{\text{fMet}}$ to fMet-tRNA$^{\text{fMet}}$ in eubacteria, is likely to be a related enzyme. A hydrolytic activity has also been identified that destroys mischarged Val-tRNA$^{\text{Ile}}$ (Schimmel, P. et al. (1998) FASEB J. 12:1599-1609). One likely scenario for the evolution of Glu-AdT in primitive life forms is the absence of a specific glutaminyl-tRNA synthetase (GlnRS), requiring an alternative pathway for the synthesis of Gln-tRNA$^{\text{Gln}}$. In fact, deletion of the Glu-AdT operon in Gram positive bacteria is lethal (Curnow, A.W. et al. (1997) Proc. Natl. Acad. Sci. U.S.A. 94:11819-11826). The existence of GluRS activity in other organisms has been inferred by the high degree of conservation in translation machinery in nature; however, GluRS has not been identified in all organisms, including Homo sapiens. Such an enzyme would be responsible for ensuring translational fidelity and reducing the synthesis of defective polypeptides,

In addition to their function in protein synthesis, specific aminoacyl tRNA synthetases also play roles in cellular fidelity, RNA splicing, RNA trafficking, apoptosis, and transcriptional and translational regulation. For example, human tyrosyl-tRNA synthetase can be proteolytically cleaved into two fragments with distinct cytokine activities. The carboxy-terminal domain exhibits monocyte and leukocyte chemotaxis activity as well as stimulating production of myeloperoxidase, tumor necrosis factor-a, and tissue factor. The N-terminal domain binds to the interleukin-8 type A receptor and functions as an interleukin-8-like cytokine. Human tyrosyl-tRNA synthetase is secreted from apoptotic tumor cells and may accelerate apoptosis (Wakasugi, K., and Schimmel, P. (1999) Science 284:147-151). Mitochondrial Neurospora crassa TyrRS and S. cerevisiae LeuRS are essential factors for certain group I intron splicing activities, and human mitochondrial LeuRS can substitute for the yeast LeuRS in a yeast null strain. Certain bacterial aaRSs are involved in regulating their own transcription or translation (Martinis, supra). Several aaRSs are able to synthesize diadenosine oligophosphates, a class of signalling molecules with roles in cell proliferation, differentiation, and apoptosis (Kisselev, L.L et al. (1998) FEBS Lett. 427:157-163; Vartanian, A. et al. (1999) FEBS Lett. 456:175-180).

Autoantibodies against aminoacyl-tRNAs are generated by patients with autoimmune diseases such as rheumatic arthritis, dermatomyositis and polymyositis, and correlate strongly with complicating

interstitial lung disease (ILD) (Freist, W. et al. (1999) Biol. Chem. 380:623-646; Freist, W. et al. (1996) Biol. Chem. Hoppe Seyler 377:343-356). These antibodies appear to be generated in response to viral infection, and coxsackie virus has been used to induce experimental viral myositis in animals.

Comparison of aaRS structures between humans and pathogens has been useful in the design of novel antibiotics (Schimmel, supra). Genetically engineered aaRSs have been utilized to allow site-specific incorporation of unnatural amino acids into proteins in vivo (Liu, D.R. et al. (1997) Proc. Natl. Acad. Sci. USA 94:10092-10097).

tRNA Modifications

The modified ribonucleoside, pseudouridine (y), is present ubiquitously in the anticodon regions of transfer RNAs (tRNAs), large and small ribosomal RNAs (rRNAs), and small nuclear RNAs (snRNAs). y is the most common of the modified nucleosides (i.e., other than G, A, U, and C) present in tRNAs. Only a few yeast tRNAs that are not involved in protein synthesis do not contain y (Cortese, R. et al. (1974) J. Biol. Chem. 249:1103-1108). The enzyme responsible for the conversion of uridine to y, pseudouridine synthase (pseudouridylate synthase), was first isolated from Salmonella typhimurium (Arena, F. et al. (1978) Nuc. Acids Res. 5:4523-4536). The enzyme has since been isolated from a number of mammals, including steer and mice (Green, C.J. et al. (1982) J. Biol. Chem. 257:3045-52 and Chen, J. and Patton, J.R. (1999) RNA 5:409-419). tRNA pseudouridine synthases have been the most extensively studied members of the family. They require a thiol donor (e.g., cysteine) and a monovalent cation (e.g., ammonia or potassium) for optimal activity. Additional cofactors or high energy molecules (e.g., ATP or GTP) are not required (Green, *supra*). Other eukaryotic pseudouridine synthases have been identified that appear to be specific for rRNA (reviewed in Smith, C.M. and Steitz, J.A. (1997) Cell 89:669-672) and a dual-specificity enzyme has been identified that uses both tRNA and rRNA substrates (Wrzesinski, J. et al. (1995) RNA 1: 437-448). The absence of y in the anticodon loop of tRNAs results in reduced growth in both bacteria (Singer, C.E. et al. (1972) Nature New Biol. 238:72-74) and yeast (Lecointe, F. (1998) 273:1316-1323), although the genetic defect is not lethal.

Another ribonucleoside modification that occurs primarily in eukaryotic cells is the conversion of guanosine to N^2,N^2-dimethylguanosine (m^2_2G) at position 26 or 10 at the base of the D-stem of cytosolic and mitochondrial tRNAs. This posttranscriptional modification is believed to stabilize tRNA structure by preventing the formation of alternative tRNA secondary and tertiary structures. Yeast $tRNA^{Asp}$ is unusual in that it does not contain this modification. The modification does not occur in eubacteria, presumably because the structure of tRNAs in these cells and organelles is sequence constrained and does not require posttranscriptional modification to prevent the formation of

alternative structures (Steinberg, S. and Cedergren, R. (1995) RNA 1:886-891, and references within). The enzyme responsible for the conversion of guanosine to $m^2{}_2G$ is a 63 kDa *S*-adenosylmethionine (SAM)-dependent tRNA N^2,N^2-dimethyl-guanosine methyltransferase (also referred to as the *TRM1* gene product and herein referred to as TRM) (Edqvist, J. (1995) Biochimie 77:54-61). The enzyme localizes to both the nucleus and the mitochondria (Li, J-M. et al. (1989) J. Cell Biol. 109:1411-1419). Based on studies with TRM from Xenopus laevis, there appears to be a requirement for base pairing at positions C11-G24 and G10-C25 immediately preceding the G26 to be modified, with other structural features of the tRNA also being required for the proper presentation of the G26 substrate (Edqvist. J. et al. (1992) Nuc. Acids Res. 20:6575-81). Studies in yeast suggest that cells carrying a weak ochre tRNA suppressor (sup3-i) are unable to suppress translation termination in the absence of TRM activity, suggesting a role for TRM in modifying the frequency of suppression in eukaryotic cells (Niederberger, C. et al. (1999) FEBS Lett. 464:67-70), in addition to the more general function of ensuring the proper three-dimensional structures for tRNA.

Translation Initiation

Initiation of translation can be divided into three stages. The first stage brings an initiator transfer RNA (Met-$tRNA_f$) together with the 40S ribosomal subunit to form the 43S preinitiation complex. The second stage binds the 43S preinitiation complex to the mRNA, followed by migration of the complex to the correct AUG initiation codon. The third stage brings the 60S ribosomal subunit to the 40S subunit to generate an 80S ribosome at the inititation codon. Regulation of translation primarily involves the first and second stage in the initiation process (V.M. Pain (1996) Eur. J. Biochem. 236:747-771).

Several initiation factors, many of which contain multiple subunits, are involved in bringing an initiator tRNA and the 40S ribosomal subunit together. eIF2, a guanine nucleotide binding protein, recruits the initiator tRNA to the 40S ribosomal subunit. Only when eIF2 is bound to GTP does it associate with the initiator tRNA. eIF2B, a guanine nucleotide exchange protein, is responsible for converting eIF2 from the GDP-bound inactive form to the GTP-bound active form. Two other factors, eIF1A and eIF3 bind and stabilize the 40S subunit by interacting with the 18S ribosomal RNA and specific ribosomal structural proteins. eIF3 is also involved in association of the 40S ribosomal subunit with mRNA. The Met-$tRNA_f$, eIF1A, eIF3, and 40S ribosomal subunit together make up the 43S preinitiation complex (Pain, supra).

Additional factors are required for binding of the 43S preinitiation complex to an mRNA molecule, and the process is regulated at several levels. eIF4F is a complex consisting of three proteins: eIF4E, eIF4A, and eIF4G. eIF4E recognizes and binds to the mRNA 5'-terminal m^7GTP

cap, eIF4A is a bidirectional RNA-dependent helicase, and eIF4G is a scaffolding polypeptide. eIF4G has three binding domains. The N-terminal third of eIF4G interacts with eIF4E, the central third interacts with eIF4A, and the C-terminal third interacts with eIF3 bound to the 43S preinitiation complex. Thus, eIF4G acts as a bridge between the 40S ribosomal subunit and the mRNA (M.W. Hentze (1997) Science 275:500-501).

The ability of eIF4F to initiate binding of the 43S preinitiation complex is regulated by structural features of the mRNA. The mRNA molecule has an untranslated region (UTR) between the 5' cap and the AUG start codon. In some mRNAs this region forms secondary structures that impede binding of the 43S preinitiation complex. The helicase activity of eIF4A is thought to function in removing this secondary structure to facilitate binding of the 43S preinitiation complex (Pain, supra).

Translation Elongation

Elongation is the process whereby additional amino acids are joined to the initiator methionine to form the complete polypeptide chain. The elongation factors EF1 a, EF1 b g, and EF2 are involved in elongating the polypeptide chain following initiation. EF1 a is a GTP-binding protein. In EF1 a's GTP-bound form, it brings an aminoacyl-tRNA to the ribosome's A site. The amino acid attached to the newly arrived aminoacyl-tRNA forms a peptide bond with the initiatior methionine. The GTP on EF1 a is hydrolyzed to GDP, and EF1 a -GDP dissociates from the ribosome. EF1 b g binds EF1 a - GDP and induces the dissociation of GDP from EF1 a, allowing EF1 a to bind GTP and a new cycle to begin.

As subsequent aminoacyl-tRNAs are brought to the ribosome, EF-G, another GTP-binding protein, catalyzes the translocation of tRNAs from the A site to the P site and finally to the E site of the ribosome. This allows the ribosome and the mRNA to remain attached during translation.

Translation Termination

The release factor eRF carries out termination of translation. eRF recognizes stop codons in the mRNA, leading to the release of the polypeptide chain from the ribosome.

Expression Profiling

Array technology can provide a simple way to explore the expression of a single polymorphic gene or the expression profile of a large number of related or unrelated genes. When the expression of a single gene is examined, arrays are employed to detect the expression of a specific gene or its variants. When an expression profile is examined, arrays provide a platform for identifying genes that are tissue specific, are affected by a substance being tested in a toxicology assay, are part of a signaling cascade, carry out housekeeping functions, or are specifically related to a particular genetic predisposition, condition, disease, or disorder.

Expression

Tumor necrosis factor α is a pleiotropic cytokine that mediates immune regulation and inflammatory responses. TNF-α-related cytokines generate partially overlapping cellular responses, including differentiation, proliferation, nuclear factor-κB (NF-κB) activation, and cell death, by triggering the aggregation of receptor monomers (Smith, C.A. et al. (1994) Cell 76:959-962). The cellular responses triggered by TNF-α are initiated through its interaction with distinct cell surface receptors (TNFRs). NF-κB is a transcription factor with a pivotal role in inducing genes involved in physiological processes as well as in the response to injury and infection. Activation of NF-κB involves the phosphorylation and subsequent degradation of an inhibitory protein, IKB, and many of the proximal kinases and adaptor molecules involved in this process have been elucidated. Additionally, the NF-κB activation pathway from cell membrane to nucleus for IL-1 and TNF-α is now understood (Bowie, A. and L.A. O'Neill (2000) Biochem. Pharmacol. 59:13-23).

Treatment of confluent cultures of vascular smooth muscle cells (SMCs) with TNF-α suppresses the incorporation of [^{3}H]proline into both collagenase-digestible proteins (CDP) and noncollagenous proteins (NCP). Such suppression by TNF-α is not observed in confluent bovine aortic endothelial cells and human fibroblastic IMR-90 cells. TNF-α decreases the relative proportion of collagen types IV and V suggesting that TNF-α modulates collagen synthesis by SMCs depending on their cell density and therefore may modify formation of atherosclerotic lesions (Hiraga, S. et al. (2000) Life Sci. 66:235-244).

Human aortic endothelial cells (HAECs) are primary cells derived from the endothelium of a human aorta. Human iliac artery endothelial cells (HIAECs) are primary cells derived from the endothelium of an iliac artery. Human umbilical vein endothelial cells (HUVECs) are primary cells derived from the endothelium of an umbilical vein. Primary human endothelial cell lines have been used as an experimental model for investigating in vitro the role of the endothelium in human vascular biology. Activation of the vascular endothelium is considered to be a central event in a wide range of both physiological and pathophysiological processes, such as vascular tone regulation, coagulation and thrombosis, atherosclerosis, and inflammation.

Thus, vascular tissue genes differentially expressed during treatment of HAEC, HIAEC, and HUVEC cell cultures with TNFα may reasonably be expected to be markers of the atherosclerotic process.

The discovery of new molecules for disease detection and treatment, and the polynucleotides encoding them, satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of cell proliferative, autoimmune/inflammatory, developmental,

and neurological disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of molecules for disease detection and treatment.

SUMMARY OF THE INVENTION

The invention features purified polypeptides, molecules for disease detection and treatment, referred to collectively as "MDDT" and individually as "MDDT-1," "MDDT-2," "MDDT-3," "MDDT-4," "MDDT-5," "MDDT-6," "MDDT-7," "MDDT-8," "MDDT-9," "MDDT-10," "MDDT-11," "MDDT-12," "MDDT-13," "MDDT-14," "MDDT-15," "MDDT-16," "MDDT-17," "MDDT-18," "MDDT-19," "MDDT-20," "MDDT-21," "MDDT-22," and "MDDT-23." In one aspect, the invention provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-23.

The invention further provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-23. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:24-46.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, and d) an immunogenic fragment of a polypeptide having an amino

acid sequence selected from the group consisting of SEQ ID NO:1-23. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23.

The invention further provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:24-46, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:24-46, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:24-46, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:24-46,

c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:24-46, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:24-46, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-23. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional MDDT, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected

from the group consisting of SEQ ID NO:1-23, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional MDDT, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional MDDT, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the

activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:24-46, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, b) detecting altered expression of the target polynucleotide, and c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:24-46, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:24-46, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:24-46, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:24-46, iii) a polynucleotide

complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the present invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog, and the PROTEOME database identification numbers and annotations of PROTEOME database homologs, for polypeptides of the invention. The probability scores for the matches between each polypeptide and its homolog(s) are also shown.

Table 3 shows structural features of polypeptide sequences of the invention, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble polynucleotide sequences of the invention, along with selected fragments of the polynucleotide sequences.

Table 5 shows the representative cDNA library for polynucleotides of the invention.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

Table 8 shows single nucleotide polymorphisms found in polynucleotide sequences of the invention, along with allele frequencies in different human populations.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing

particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"MDDT" refers to the amino acid sequences of substantially purified MDDT obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of MDDT. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of MDDT either by directly interacting with MDDT or by acting on components of the biological pathway in which MDDT participates.

An "allelic variant" is an alternative form of the gene encoding MDDT. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding MDDT include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as MDDT or a polypeptide with at least one functional characteristic of MDDT. Included within this definition are

polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding MDDT, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding MDDT. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent MDDT. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of MDDT is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of MDDT. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of MDDT either by directly interacting with MDDT or by acting on components of the biological pathway in which MDDT participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, $F(ab')_2$, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind MDDT polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and

keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "aptamer" refers to a nucleic acid or oligonucleotide molecule that binds to a specific molecular target. Aptamers are derived from an in vitro evolutionary process (e.g., SELEX (Systematic Evolution of Ligands by EXponential Enrichment), described in U.S. Patent No. 5,270,163), which selects for target-specific aptamer sequences from large combinatorial libraries. Aptamer compositions may be double-stranded or single-stranded, and may include deoxyribonucleotides, ribonucleotides, nucleotide derivatives, or other nucleotide-like molecules. The nucleotide components of an aptamer may have modified sugar groups (e.g., the 2'-OH group of a ribonucleotide may be replaced by 2'-F or 2'-NH_2), which may improve a desired property, e.g., resistance to nucleases or longer lifetime in blood. Aptamers may be conjugated to other molecules, e.g., a high molecular weight carrier to slow clearance of the aptamer from the circulatory system. Aptamers may be specifically cross-linked to their cognate ligands, e.g., by photo-activation of a cross-linker. (See, e.g., Brody, E.N. and L. Gold (2000) J. Biotechnol. 74:5-13.)

The term "intramer" refers to an aptamer which is expressed in vivo. For example, a vaccinia virus-based RNA expression system has been used to express specific RNA aptamers at high levels in the cytoplasm of leukocytes (Blind, M. et al. (1999) Proc. Natl Acad. Sci. USA 96:3606-3610).

The term "spiegelmer" refers to an aptamer which includes L-DNA, L-RNA, or other left-handed nucleotide derivatives or nucleotide-like molecules. Aptamers containing left-handed nucleotides are resistant to degradation by naturally occurring enzymes, which normally act on substrates containing right-handed nucleotides.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once

introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic MDDT, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding MDDT or fragments of MDDT may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

Original Residue	Conservative Substitution
Ala	Gly, Ser
Arg	His, Lys
Asn	Asp, Gln, His
Asp	Asn, Glu
Cys	Ala, Ser
Gln	Asn, Glu, His
Glu	Asp, Gln, His
Gly	Ala
His	Asn, Arg, Gln, Glu
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile
Phe	His, Met, Leu, Trp, Tyr
Ser	Cys, Thr
Thr	Ser, Val
Trp	Phe, Tyr
Tyr	His, Phe, Trp
Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

"Differential expression" refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

"Exon shuffling" refers to the recombination of different coding regions (exons). Since an

exon may represent a structural or functional domain of the encoded protein, new proteins may be assembled through the novel reassortment of stable substructures, thus allowing acceleration of the evolution of new protein functions.

A "fragment" is a unique portion of MDDT or the polynucleotide encoding MDDT which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:24-46 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:24-46, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:24-46 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:24-46 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:24-46 and the region of SEQ ID NO:24-46 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-23 is encoded by a fragment of SEQ ID NO:24-46. A fragment of SEQ ID NO:1-23 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-23. For example, a fragment of SEQ ID NO:1-23 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-23. The precise length of a fragment of SEQ ID NO:1-23 and the region of SEQ ID NO:1-23 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A "full length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two

or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at http://www.ncbi.nlm.nih.gov/BLAST/. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at http://www.ncbi.nlm.nih.gov/gorf/bl2.html. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

Penalty for mismatch: -2

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 11

Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 μg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of

the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 μg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of MDDT which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of MDDT which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides,

polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of MDDT. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of MDDT.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an MDDT may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of MDDT.

"Probe" refers to nucleic acid sequences encoding MDDT, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous

nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially

complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing MDDT, nucleic acids encoding MDDT, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure

of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" or "expression profile" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. In one alternative, the nucleic acid can be introduced by infection with a recombinant viral vector, such as a lentiviral vector (Lois, C. et al. (2002) Science 295:868-872). The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is

directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), supra.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length of one of the polypeptides.

THE INVENTION

The invention is based on the discovery of new human molecules for disease detection and treatment (MDDT), the polynucleotides encoding MDDT, and the use of these compositions for the diagnosis, treatment, or prevention of cell proliferative, autoimmune/inflammatory, developmental, and neurological disorders, and infections.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown. Column 6 shows the Incyte ID numbers of physical, full length clones corresponding to the polypeptide and polynucleotide sequences of the invention. The full length clones encode polypeptides which have at least 95% sequence identity to the polypeptide sequences shown in column 3.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database and the PROTEOME database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (GenBank ID NO:) of the nearest GenBank homolog and the PROTEOME database identification numbers (PROTEOME ID NO:) of the nearest PROTEOME database homologs. Column 4 shows the probability scores for the matches between each polypeptide and its homolog(s). Column 5 shows the annotation of the GenBank and PROTEOME database homolog(s) along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide.. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are molecules for disease detection and treatment. For example, SEQ ID NO:1 is 42% identical, from residue M1 to residue D482, to human RO52 gene product (GenBank ID g747927) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 9.8e-97, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:1 also contains a SPRY domain, a B-box zinc finger domain, and a RING finger C3HC4 type zinc finger domain, as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:1 is a transcription factor. In another example, SEQ ID NO:9 is 86% identical, from residue M1 to residue R722, to mouse DNA binding protein DESRT (GenBank ID g9622226) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:9 also contains an ARID DNA binding domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. Data from further BLAST analyses provide further corroborative evidence that SEQ ID NO:9 is a DNA-binding protein. In a further example, SEQ ID NO:11 is 81% identical, from residue R8 to residue S86, to human HERV-E integrase (GenBank ID g2587026) as determined by the Basic Local Alignment Search Tool (BLAST). The BLAST probability score is 2.7e-32, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. Data from BLAST analyses provide further corroborative evidence that SEQ ID NO:11 is an integrase protease. In yet a further example, SEQ ID NO:16 is 98% identical, from residue M1 to residue A928, to human prostate antigen PARIS-1 (GenBank ID g12963885) as determined by the Basic Local Alignment Search Tool (BLAST). The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:16 also contains a PH domain and a TBC domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. Data from BLIMPS and BLAST analyses provide further corroborative evidence that SEQ ID NO:16 is a full-length human protein for disease detection and treatment. SEQ ID NO:2-8, SEQ ID NO:10, SEQ ID NO:12-15, and SEQ ID NO:17-23 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-23 are described in Table 7.

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Column 1 lists the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:), the corresponding Incyte polynucleotide consensus sequence number (Incyte ID) for each polynucleotide of the invention, and the length of each polynucleotide sequence in basepairs. Column 2 shows the nucleotide start (5′) and stop (3′) positions of the cDNA and/or genomic sequences used to assemble the full length polynucleotide sequences of the invention, and of fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:24-46 or that distinguish between SEQ ID NO:24-46 and related polynucleotide sequences.

The polynucleotide fragments described in Column 2 of Table 4 may refer specifically, for example, to Incyte cDNAs derived from tissue-specific cDNA libraries or from pooled cDNA libraries. Alternatively, the polynucleotide fragments described in column 2 may refer to GenBank cDNAs or ESTs which contributed to the assembly of the full length polynucleotide sequences. In addition, the polynucleotide fragments described in column 2 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation "ENST"). Alternatively, the polynucleotide fragments described in column 2 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences including the designation "NM" or "NT") or the NCBI RefSeq Protein Sequence Records (*i.e.*, those sequences including the designation "NP"). Alternatively, the polynucleotide fragments described in column 2 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, a polynucleotide sequence identified as FL_$XXXXXX$_N_1_N_2_$YYYYY$_N_3_N_4 represents a "stitched" sequence in which *XXXXXX* is the identification number of the cluster of sequences to which the algorithm was applied, and *YYYYY* is the number of the prediction generated by the algorithm, and $N_{1,2,3...}$, if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the polynucleotide fragments in column 2 may refer to assemblages of exons brought together by an "exon-stretching" algorithm. For example, a polynucleotide sequence identified as FL*XXXXXX*_*gAAAAA*_*gBBBBB*_1_*N* is a "stretched" sequence, with *XXXXXX* being the Incyte project identification number, *gAAAAA* being the GenBank identification number of the human genomic sequence to which the "exon-stretching" algorithm was applied, *gBBBBB* being the GenBank identification number or NCBI RefSeq identification number of the nearest GenBank protein homolog, and *N* referring to specific exons (See Example V). In instances where a RefSeq sequence was used

as a protein homolog for the "exon-stretching" algorithm, a RefSeq identifier (denoted by "NM," "NP," or "NT") may be used in place of the GenBank identifier (*i.e.*, *gBBBBB*).

Alternatively, a prefix identifies component sequences that were hand-edited, predicted from genomic DNA sequences, or derived from a combination of sequence analysis methods. The following Table lists examples of component sequence prefixes and corresponding sequence analysis methods associated with the prefixes (see Example IV and Example V).

Prefix	Type of analysis and/or examples of programs
GNN, GFG, ENST	Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK).
GBI	Hand-edited analysis of genomic sequences.
FL	Stitched or stretched genomic sequences (see Example V).
INCY	Full length transcript and exon prediction from mapping of EST sequences to the genome. Genomic location and EST composition data are combined to predict the exons and resulting transcript.

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in Table 4 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotide sequences. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

Table 8 shows single nucleotide polymorphisms (SNPs) found in polynucleotide sequences of the invention, along with allele frequencies in different human populations. Columns 1 and 2 show the polynucleotide sequence identification number (SEQ ID NO:) and the corresponding Incyte project identification number (PID) for polynucleotides of the invention. Column 3 shows the Incyte identification number for the EST in which the SNP was detected (EST ID), and column 4 shows the identification number for the SNP (SNP ID). Column 5 shows the position within the EST sequence at which the SNP is located (EST SNP), and column 6 shows the position of the SNP within the full-length polynucleotide sequence (CB1 SNP). Column 7 shows the allele found in the EST sequence. Columns 8 and 9 show the two alleles found at the SNP site. Column 10 shows the amino acid

encoded by the codon including the SNP site, based upon the allele found in the EST. Columns 11-14 show the frequency of allele 1 in four different human populations. An entry of n/d (not detected) indicates that the frequency of allele 1 in the population was too low to be detected, while n/a (not available) indicates that the allele frequency was not determined for the population.

The invention also encompasses MDDT variants. A preferred MDDT variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the MDDT amino acid sequence, and which contains at least one functional or structural characteristic of MDDT.

The invention also encompasses polynucleotides which encode MDDT. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:24-46, which encodes MDDT. The polynucleotide sequences of SEQ ID NO:24-46, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding MDDT. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding MDDT. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:24-46 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:24-46. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of MDDT.

In addition, or in the alternative, a polynucleotide variant of the invention is a splice variant of a polynucleotide sequence encoding MDDT. A splice variant may have portions which have significant sequence identity to the polynucleotide sequence encoding MDDT, but will generally have a greater or lesser number of polynucleotides due to additions or deletions of blocks of sequence arising from alternate splicing of exons during mRNA processing. A splice variant may have less than about 70%, or alternatively less than about 60%, or alternatively less than about 50% polynucleotide sequence identity to the polynucleotide sequence encoding MDDT over its entire length; however, portions of the splice variant will have at least about 70%, or alternatively at least about 85%, or alternatively at least about 95%, or alternatively 100% polynucleotide sequence identity to portions of the polynucleotide sequence encoding MDDT. For example, a polynucleotide comprising a sequence of

SEQ ID NO:25 is a splice variant of a polynucleotide comprising a sequence of SEQ ID NO:45, and a polynucleotide comprising a sequence of SEQ ID NO:36 is a splice variant of a polynucleotide comprising a sequence of SEQ ID NO:46. Any one of the splice variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of MDDT.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding MDDT, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring MDDT, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode MDDT and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring MDDT under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding MDDT or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding MDDT and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode MDDT and MDDT derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding MDDT or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:24-46 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in

"Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding MDDT may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National

Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode MDDT may be cloned in recombinant DNA molecules that direct expression of MDDT, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express MDDT.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter MDDT-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent No. 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve

the biological properties of MDDT, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding MDDT may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively, MDDT itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of MDDT, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, supra, pp. 28-53.)

In order to express a biologically active MDDT, the nucleotide sequences encoding MDDT or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding MDDT. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding MDDT. Such signals

include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding MDDT and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding MDDT and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding MDDT. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO J. 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; and Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5(6):350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90(13):6340-6344; Buller, R.M. et al. (1985) Nature 317(6040):813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31(3):219-226; and Verma, I.M. and N. Somia (1997) Nature 389:239-242.) The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding MDDT. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding MDDT can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding MDDT into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of MDDT are needed, e.g. for the production of antibodies, vectors which direct high level expression of MDDT may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of MDDT. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; and Scorer, C.A. et al. (1994) Bio/Technology 12:181-184.)

Plant systems may also be used for expression of MDDT. Transcription of sequences encoding MDDT may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding MDDT may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses MDDT in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc.

Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of MDDT in cell lines is preferred. For example, sequences encoding MDDT can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk⁻* and *apr⁻* cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), ß glucuronidase and its substrate ß-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest

is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding MDDT is inserted within a marker gene sequence, transformed cells containing sequences encoding MDDT can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding MDDT under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding MDDT and that express MDDT may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of MDDT using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on MDDT is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding MDDT include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding MDDT, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding MDDT may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode MDDT may be designed to contain signal sequences which direct secretion of MDDT through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding MDDT may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric MDDT protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of MDDT activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the MDDT encoding sequence and the heterologous protein sequence, so that MDDT may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled MDDT may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ^{35}S-methionine.

MDDT of the present invention or fragments thereof may be used to screen for compounds that specifically bind to MDDT. At least one and up to a plurality of test compounds may be screened for specific binding to MDDT. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of MDDT, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which MDDT binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express MDDT, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing MDDT or cell membrane fractions which contain MDDT are then contacted with a test compound and binding, stimulation, or inhibition of activity of either MDDT or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with MDDT, either in solution or affixed to a solid support, and detecting the binding of MDDT to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

MDDT of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of MDDT. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for MDDT activity, wherein MDDT is combined with at least one test compound, and the activity of MDDT in the presence of a test compound is compared with the activity of MDDT in the absence of the test compound. A change in the activity of MDDT in the presence of the test compound is indicative of a

compound that modulates the activity of MDDT. Alternatively, a test compound is combined with an in vitro or cell-free system comprising MDDT under conditions suitable for MDDT activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of MDDT may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding MDDT or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding MDDT may also be manipulated in vitro in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding MDDT can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding MDDT is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress MDDT, e.g., by secreting MDDT in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of MDDT and molecules for disease detection and treatment. In addition, examples of tissues and cell lines expressing MDDT are vascular smooth muscle cells, human aortic endothelial cells, human iliac artery endothelial cells, and human umbilical vein endothelial cells, and also can be found in Table 6. Therefore, MDDT appears to play a role in cell proliferative, autoimmune/inflammatory, developmental, and neurological disorders, and infections. In the treatment of disorders associated with increased MDDT expression or activity, it is desirable to decrease the expression or activity of MDDT. In the treatment of disorders associated with decreased MDDT expression or activity, it is desirable to increase the expression or activity of MDDT.

Therefore, in one embodiment, MDDT or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of MDDT. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, a cancer of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; an autoimmune/inflammatory disorder such as inflammation, actinic keratosis, acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, arteriosclerosis, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, bronchitis, bursitis, cholecystitis, cirrhosis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, paroxysmal nocturnal hemoglobinuria, hepatitis, hypereosinophilia, irritable bowel syndrome, episodic lymphopenia with lymphocytotoxins, mixed connective tissue disease (MCTD), multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, myelofibrosis, osteoarthritis, osteoporosis, pancreatitis, polycythemia vera, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, primary thrombocythemia, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), episodic lymphopenia with

lymphocytotoxins, complications of cancer, hemodialysis, and extracorporeal circulation, trauma, and hematopoietic cancer including lymphoma, leukemia, and myeloma; a developmental disorder such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorder of the central nervous system, cerebral palsy, a neuroskeletal disorder, an autonomic nervous system disorder, a cranial nerve disorder, a spinal cord disease, muscular dystrophy and other neuromuscular disorder, a peripheral nervous system disorder, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathy, myasthenia gravis, periodic paralysis, a mental disorder including mood, anxiety, and schizophrenic disorder, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, and Tourette's disorder; and an infection, such as those caused by a viral agent classified as adenovirus (acute respiratory disease, pneumonia), arenavirus (lymphocytic choriomeningitis), bunyavirus (Hantavirus), calicivirus, coronavirus (pneumonia, chronic bronchitis), filovirus, hepadnavirus (hepatitis), herpesvirus (herpes simplex virus, varicella-zoster virus, Epstein-Barr virus, cytomegalovirus), flavivirus (yellow fever), orthomyxovirus (influenza), parvovirus, papovavirus or papillomaviruse (cancer), paramyxovirus (measles, mumps), picornavirus (rhinovirus, poliovirus, coxsackie-virus), polyomaviruse (BK virus, JC virus), poxviruse (smallpox), reoviru (Colorado tick fever), retroviruse (human immunodeficiency virus, human T lymphotropic virus), rhabdoviruse (rabies), rotaviruse (gastroenteritis), and togaviruse

(encephalitis, rubella); an infection caused by a bacterial agent classified as pneumococcus, staphylococcus, streptococcus, bacillus, corynebacterium, clostridium, meningococcus, gonococcus, listeria, moraxella, kingella, haemophilus, legionella, bordetella, gram-negative enterobacterium including shigella, salmonella, or campylobacter, pseudomonas, vibrio, brucella, francisella, yersinia, bartonella, norcardium, actinomyces, mycobacterium, spirochaetale, rickettsia, chlamydia, or mycoplasma; an infection caused by a fungal agent classified as aspergillus, blastomyces, dermatophytes, cryptococcus, coccidioides, malasezzia, histoplasma, or other mycosis-causing fungal agent; and an infection caused by a parasite classified as plasmodium or malaria-causing, parasitic entamoeba, leishmania, trypanosoma, toxoplasma, pneumocystis carinii, intestinal protozoa such as giardia, trichomonas, tissue nematode such as trichinella, intestinal nematode such as ascaris, lymphatic filarial nematode, trematode such as schistosoma, and cestode such as tapeworm.

In another embodiment, a vector capable of expressing MDDT or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of MDDT including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified MDDT in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of MDDT including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of MDDT may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of MDDT including, but not limited to, those listed above.

In a further embodiment, an antagonist of MDDT may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of MDDT. Examples of such disorders include, but are not limited to, those cell proliferative, autoimmune/inflammatory, developmental, and neurological disorders, and infections described above. In one aspect, an antibody which specifically binds MDDT may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express MDDT.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding MDDT may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of MDDT including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made

by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of MDDT may be produced using methods which are generally known in the art. In particular, purified MDDT may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind MDDT. Antibodies to MDDT may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use. Single chain antibodies (e.g., from camels or llamas) may be potent enzyme inhibitors and may have advantages in the design of peptide mimetics, and in the development of immuno-adsorbents and biosensors (Muyldermans, S. (2001) J. Biotechnol. 74:277-302).

For the production of antibodies, various hosts including goats, rabbits, rats, mice, camels, dromedaries, llamas, humans, and others may be immunized by injection with MDDT or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to MDDT have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of MDDT amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to MDDT may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and

Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce MDDT-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for MDDT may also be generated. For example, such fragments include, but are not limited to, $F(ab')_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between MDDT and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering MDDT epitopes is generally used, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for MDDT. Affinity is expressed as an association constant, K_a, which is defined as the molar concentration of MDDT-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple MDDT epitopes, represents the average affinity, or avidity, of the antibodies for MDDT. The

K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular MDDT epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the MDDT-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of MDDT, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of MDDT-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding MDDT, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding MDDT. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding MDDT. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) J. Allergy Clin. Immunol. 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) Blood 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) Pharmacol. Ther. 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other

systems known in the art. (See, e.g., Rossi, J.J. (1995) Br. Med. Bull. 51(1):217-225; Boado, R.J. et al. (1998) J. Pharm. Sci. 87(11):1308-1315; and Morris, M.C. et al. (1997) Nucleic Acids Res. 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding MDDT may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) Science 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassamias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in MDDT expression or regulation causes disease, the expression of MDDT from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in MDDT are treated by constructing mammalian expression vectors encoding MDDT and introducing these vectors by mechanical means into MDDT-deficient cells. Mechanical transfer technologies for use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of MDDT include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX, PCR2-TOPOTA vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA),

and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). MDDT may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β-actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and H.M. Blau, supra)), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding MDDT from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to MDDT expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding MDDT under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent No. 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., $CD4^+$ T-cells), and the

return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding MDDT to cells which have one or more genetic abnormalities with respect to the expression of MDDT. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent No. 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) Annu. Rev. Nutr. 19:511-544 and Verma, I.M. and N. Somia (1997) Nature 18:389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding MDDT to target cells which have one or more genetic abnormalities with respect to the expression of MDDT. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing MDDT to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res. 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent No. 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent No. 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) J. Virol. 73:519-532 and Xu, H. et al. (1994) Dev. Biol. 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding MDDT to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for MDDT into the alphavirus genome in place of the capsid-coding region results in the production of a large number of MDDT-coding RNAs and the synthesis of high levels of MDDT in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of MDDT into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding MDDT.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding MDDT. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding MDDT. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased MDDT expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding MDDT may be therapeutically useful, and in the treatment of disorders associated with decreased MDDT expression or activity, a compound which specifically promotes expression of the

polynucleotide encoding MDDT may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding MDDT is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an in vitro cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding MDDT are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding MDDT. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a Schizosaccharomyces pombe gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and

monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of MDDT, antibodies to MDDT, and mimetics, agonists, antagonists, or inhibitors of MDDT.

The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising MDDT or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, MDDT or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for

administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example MDDT or fragments thereof, antibodies of MDDT, and agonists, antagonists or inhibitors of MDDT, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind MDDT may be used for the diagnosis of disorders characterized by expression of MDDT, or in assays to monitor patients being treated with MDDT or agonists, antagonists, or inhibitors of MDDT. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for MDDT include methods which utilize the antibody and a label to detect MDDT in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification,

and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring MDDT, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of MDDT expression. Normal or standard values for MDDT expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to MDDT under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of MDDT expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding MDDT may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of MDDT may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of MDDT, and to monitor regulation of MDDT levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding MDDT or closely related molecules may be used to identify nucleic acid sequences which encode MDDT. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding MDDT, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the MDDT encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:24-46 or from genomic sequences including promoters, enhancers, and introns of the MDDT gene.

Means for producing specific hybridization probes for DNAs encoding MDDT include the cloning of polynucleotide sequences encoding MDDT or MDDT derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes <u>in vitro</u> by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels,

such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding MDDT may be used for the diagnosis of disorders associated with expression of MDDT. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, a cancer of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; an autoimmune/inflammatory disorder such as inflammation, actinic keratosis, acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, arteriosclerosis, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, bronchitis, bursitis, cholecystitis, cirrhosis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, paroxysmal nocturnal hemoglobinuria, hepatitis, hypereosinophilia, irritable bowel syndrome, episodic lymphopenia with lymphocytotoxins, mixed connective tissue disease (MCTD), multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, myelofibrosis, osteoarthritis, osteoporosis, pancreatitis, polycythemia vera, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, primary thrombocythemia, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), episodic lymphopenia with lymphocytotoxins, complications of cancer, hemodialysis, and extracorporeal circulation, trauma, and hematopoietic cancer including lymphoma, leukemia, and myeloma; a developmental disorder such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; a neurological disorder such as epilepsy, ischemic cerebrovascular disease,

stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorder of the central nervous system, cerebral palsy, a neuroskeletal disorder, an autonomic nervous system disorder, a cranial nerve disorder, a spinal cord disease, muscular dystrophy and other neuromuscular disorder, a peripheral nervous system disorder, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathy, myasthenia gravis, periodic paralysis, a mental disorder including mood, anxiety, and schizophrenic disorder, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, and Tourette's disorder; and an infection, such as those caused by a viral agent classified as adenovirus (acute respiratory disease, pneumonia), arenavirus (lymphocytic choriomeningitis), bunyavirus (Hantavirus), calicivirus, coronavirus (pneumonia, chronic bronchitis), filovirus, hepadnavirus (hepatitis), herpesvirus (herpes simplex virus, varicella-zoster virus, Epstein-Barr virus, cytomegalovirus), flavivirus (yellow fever), orthomyxovirus (influenza), parvovirus, papovavirus or papillomaviruse (cancer), paramyxovirus (measles, mumps), picornavirus (rhinovirus, poliovirus, coxsackie-virus), polyomaviruse (BK virus, JC virus), poxviruse (smallpox), reoviru (Colorado tick fever), retroviruse (human immunodeficiency virus, human T lymphotropic virus), rhabdoviruse (rabies), rotaviruse (gastroenteritis), and togaviruse (encephalitis, rubella); an infection caused by a bacterial agent classified as pneumococcus, staphylococcus, streptococcus, bacillus, corynebacterium, clostridium, meningococcus, gonococcus, listeria, moraxella, kingella, haemophilus, legionella, bordetella, gram-negative enterobacterium including shigella, salmonella, or campylobacter, pseudomonas, vibrio, brucella, francisella, yersinia, bartonella, norcardium, actinomyces, mycobacterium, spirochaetale, rickettsia, chlamydia, or mycoplasma; an infection caused by a fungal agent classified as aspergillus, blastomyces, dermatophytes, cryptococcus, coccidioides, malasezzia, histoplasma, or other mycosis-causing fungal agent; and an infection caused by a parasite classified as plasmodium or malaria-causing, parasitic entamoeba, leishmania, trypanosoma, toxoplasma, pneumocystis carinii, intestinal protozoa such as giardia, trichomonas, tissue nematode such as trichinella, intestinal nematode such as

ascaris, lymphatic filarial nematode, trematode such as schistosoma, and cestode such as tapeworm. The polynucleotide sequences encoding MDDT may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered MDDT expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding MDDT may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding MDDT may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding MDDT in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of MDDT, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding MDDT, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ

preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding MDDT may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding MDDT, or a fragment of a polynucleotide complementary to the polynucleotide encoding MDDT, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding MDDT may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding MDDT are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSCCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

SNPs may be used to study the genetic basis of human disease. For example, at least 16 common SNPs have been associated with non-insulin-dependent diabetes mellitus. SNPs are also useful for examining differences in disease outcomes in monogenic disorders, such as cystic fibrosis, sickle cell anemia, or chronic granulomatous disease. For example, variants in the mannose-binding lectin, MBL2, have been shown to be correlated with deleterious pulmonary outcomes in cystic fibrosis. SNPs also have utility in pharmacogenomics, the identification of genetic variants that

influence a patient's response to a drug, such as life-threatening toxicity. For example, a variation in N-acetyl transferase is associated with a high incidence of peripheral neuropathy in response to the anti-tuberculosis drug isoniazid, while a variation in the core promoter of the ALOX5 gene results in diminished clinical response to treatment with an anti-asthma drug that targets the 5-lipoxygenase pathway. Analysis of the distribution of SNPs in different populations is useful for investigating genetic drift, mutation, recombination, and selection, as well as for tracing the origins of populations and their migrations. (Taylor, J.G. et al. (2001) Trends Mol. Med. 7:507-512; Kwok, P.-Y. and Z. Gu (1999) Mol. Med. Today 5:538-543; Nowotny, P. et al. (2001) Curr. Opin. Neurobiol. 11:637-641.)

Methods which may also be used to quantify the expression of MDDT include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, MDDT, fragments of MDDT, or antibodies specific for MDDT may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by

quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent No. 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with in vitro model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at http://www.niehs.nih.gov/oc/news/toxchip.htm.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated

biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, supra). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for MDDT to quantify the levels of MDDT expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111; Mendoze, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each

array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in <u>DNA Microarrays: A Practical Approach</u>, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding MDDT may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members

of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding MDDT on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, MDDT, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between MDDT and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds

having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with MDDT, or fragments thereof, and washed. Bound MDDT is then detected by methods well known in the art. Purified MDDT can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding MDDT specifically compete with a test compound for binding MDDT. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with MDDT.

In additional embodiments, the nucleotide sequences which encode MDDT may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications and publications, mentioned above and below, including U.S. Ser. No. 60/280,387, U.S. Ser. No. 60/282,335, U.S. Ser. No. 60/283,663, U.S. Ser. No. 60/285,484, U.S. Ser. No. 60/350,702, and U.S. Ser. No. 60/351,749, are expressly incorporated by reference herein.

EXAMPLES

I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA). Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), PCR2-TOPOTA plasmid (Invitrogen), PCMV-ICIS plasmid (Stratagene), pIGEN (Incyte Genomics, Palo Alto CA), pRARE (Incyte Genomics), or pINCY (Incyte Genomics), or derivatives thereof. Recombinant plasmids were transformed into competent E. coli cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5α, DH10B, or ElectroMAX DH10B from Life Technologies.

II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in

384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM; PROTEOME databases with sequences from Homo sapiens, Rattus norvegicus, Mus musculus, Caenorhabditis elegans, Saccharomyces cerevisiae, Schizosaccharomyces pombe, and Candida albicans (Incyte Genomics, Palo Alto CA); hidden Markov model (HMM)-based protein family databases such as PFAM, INCY, and TIGRFAM (Haft, D.H. et al. (2001) Nucleic Acids Res. 29:41-43); and HMM-based protein domain databases such as SMART (Schultz et al. (1998) Proc. Natl. Acad. Sci. USA 95:5857-5864; Letunic, I. et al. (2002) Nucleic Acids Res. 30:242-244). (HMM is a probabilistic approach which analyzes consensus primary structures of gene families. See, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences.

Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide of the invention may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, the PROTEOME databases, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, hidden Markov model (HMM)-based protein family databases such as PFAM, INCY, and TIGRFAM; and HMM-based protein domain databases such as SMART. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:24-46. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 2.

IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative molecules for disease detection and treatment were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (See Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94,

and Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode molecules for disease detection and treatment, the encoded polypeptides were analyzed by querying against PFAM models for molecules for disease detection and treatment. Potential molecules for disease detection and treatment were also identified by homology to Incyte cDNA sequences that had been annotated as molecules for disease detection and treatment. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

V. Assembly of Genomic Sequence Data with cDNA Sequence Data

"Stitched" Sequences

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear

along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbpri public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

"Stretched" Sequences

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

VI. Chromosomal Mapping of MDDT Encoding Polynucleotides

The sequences which were used to assemble SEQ ID NO:24-46 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:24-46 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between

chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (http://www.ncbi.nlm.nih.gov/genemap/), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel (1995) supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score x Percent Identity}}{5 \text{ x minimum } \{\text{length(Seq. 1), length(Seq. 2)}\}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79%

identity and 100% overlap.

Alternatively, polynucleotide sequences encoding MDDT are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding MDDT. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

VIII. Extension of MDDT Encoding Polynucleotides

Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+}, $(NH_4)_2SO_4$, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min;

Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 μl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 μl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 μl to 10 μl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent E. coli cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethysulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotide sequences are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides

designed for such extension, and an appropriate genomic library.

IX. Identification of Single Nucleotide Polymorphisms in MDDT Encoding Polynucleotides

Common DNA sequence variants known as single nucleotide polymorphisms (SNPs) were identified in SEQ ID NO:24-46 using the LIFESEQ database (Incyte Genomics). Sequences from the same gene were clustered together and assembled as described in Example III, allowing the identification of all sequence variants in the gene. An algorithm consisting of a series of filters was used to distinguish SNPs from other sequence variants. Preliminary filters removed the majority of basecall errors by requiring a minimum Phred quality score of 15, and removed sequence alignment errors and errors resulting from improper trimming of vector sequences, chimeras, and splice variants. An automated procedure of advanced chromosome analysis analysed the original chromatogram files in the vicinity of the putative SNP. Clone error filters used statistically generated algorithms to identify errors introduced during laboratory processing, such as those caused by reverse transcriptase, polymerase, or somatic mutation. Clustering error filters used statistically generated algorithms to identify errors resulting from clustering of close homologs or pseudogenes, or due to contamination by non-human sequences. A final set of filters removed duplicates and SNPs found in immunoglobulins or T-cell receptors.

Certain SNPs were selected for further characterization by mass spectrometry using the high throughput MASSARRAY system (Sequenom, Inc.) to analyze allele frequencies at the SNP sites in four different human populations. The Caucasian population comprised 92 individuals (46 male, 46 female), including 83 from Utah, four French, three Venezualan, and two Amish individuals. The African population comprised 194 individuals (97 male, 97 female), all African Americans. The Hispanic population comprised 324 individuals (162 male, 162 female), all Mexican Hispanic. The Asian population comprised 126 individuals (64 male, 62 female) with a reported parental breakdown of 43% Chinese, 31% Japanese, 13% Korean, 5% Vietnamese, and 8% other Asian. Allele frequencies were first analyzed in the Caucasian population; in some cases those SNPs which showed no allelic variance in this population were not further tested in the other three populations.

X. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:24-46 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of

[γ-^{32}P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

XI. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, supra.), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena (1999), supra). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorbtion and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on

the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)$^+$ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)$^+$ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/μl oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/μl RNase inhibitor, 500 μM dATP, 500 μM dGTP, 500 μM dTTP, 40 μM dCTP, 40 μM dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)$^+$ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)$^+$ RNAs are synthesized by in vitro transcription from non-coding yeast genomic DNA. After incubation at 37° C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85° C to the stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 μl 5X SSC/0.2% SDS.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 μg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in U.S. Patent No. 5,807,522, incorporated herein by reference. 1 μl of the array element DNA, at an average concentration of 100 ng/μl, is loaded into the open capillary printing element by a high-speed robotic

apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in 0.2% SDS and distilled water as before.

Hybridization

Hybridization reactions contain 9 μl of sample mixture consisting of 0.2 μg each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm^2 coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 μl of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X SSC), and dried.

Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location

to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

<u>Expression</u>

<u>TNF-α treatment of HAEC cultures</u>

HAECs were maintained in EGM-2 medium (Clonetics, San Diego CA) containing 2% FBS, recombinant hEGF (0.5 $ng.ml^{-1}$), Gentamicin (50 $\mu g.ml^{-1}$), and Amphotericin-B (50 $ng.ml^{-1}$) (as supplied by Clonetics), at 37°C in a 5% CO_2 atmosphere. In addition, hydrocortisone, VEGF, R3-IGF-1, ascorbic acid, hFGF-B, and heparin were included in the medium according to manufacturer's instruction (Clonetics). The cells were grown to 85% confluency and then treated with TNF-α (10 $ng.ml^{-1}$) for 1, 2, 4, 6, 8, 10, 24, and 48 hours. These TNF-α treated cells were compared to untreated HAECs collected at 85% confluency (t = 0 hour).

For SEQ ID NO:38, the expression of a component of this polynucleotide sequence, having Incyte clone ID 2662817, is downregulated by at least two-fold when treated with TNF-α in three primary endothelial cell lines, HAEC, HIAEC, and HUVEC. Incyte clone ID 2662817 spans nucleotides 474 through 1176 of Incyte polynucleotide 2457335CB1 (SEQ ID NO:38).

XII. Complementary Polynucleotides

Sequences complementary to the MDDT-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring MDDT. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same

procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of MDDT. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the MDDT-encoding transcript.

XIII. Expression of MDDT

Expression and purification of MDDT is achieved using bacterial or virus-based expression systems. For expression of MDDT in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express MDDT upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of MDDT in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding MDDT by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, MDDT is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from MDDT at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra,

ch. 10 and 16). Purified MDDT obtained by these methods can be used directly in the assays shown in Examples XVII, XVIII, and XIX, where applicable.

XIV. Functional Assays

MDDT function is assessed by expressing the sequences encoding MDDT at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of MDDT on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding MDDT and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding MDDT and other genes of interest can be analyzed by northern analysis or microarray techniques.

XV. Production of MDDT Specific Antibodies

MDDT substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize animals (e.g., rabbits, mice, etc.) and to produce antibodies using standard protocols.

Alternatively, the MDDT amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using FMOC chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-MDDT activity by, for example, binding the peptide or MDDT to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XVI. Purification of Naturally Occurring MDDT Using Specific Antibodies

Naturally occurring or recombinant MDDT is substantially purified by immunoaffinity chromatography using antibodies specific for MDDT. An immunoaffinity column is constructed by covalently coupling anti-MDDT antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing MDDT are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of MDDT (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/MDDT binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and MDDT is collected.

XVII. Identification of Molecules Which Interact with MDDT

MDDT, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent. (See, e.g., Bolton, A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled MDDT, washed, and any wells with labeled MDDT complex are assayed. Data obtained using different concentrations of MDDT are used to calculate values for the number, affinity, and association of MDDT with the

candidate molecules.

Alternatively, molecules interacting with MDDT are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989) Nature 340:245-246, or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

MDDT may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

XVIII. Demonstration of MDDT Activity

A microtubule motility assay for MDDT measures motor protein activity. In this assay, recombinant MDDT is immobilized onto a glass slide or similar substrate. Taxol-stabilized bovine brain microtubules (commercially available) in a solution containing ATP and cytosolic extract are perfused onto the slide. Movement of microtubules as driven by MDDT motor activity can be visualized and quantified using video-enhanced light microscopy and image analysis techniques. MDDT activity is directly proportional to the frequency and velocity of microtubule movement.

Alternatively, an assay for MDDT activity measures the formation of protein filaments in vitro. A solution of MDDT at a concentration greater than the "critical concentration" for polymer assembly is applied to carbon-coated grids. Appropriate nucleation sites may be supplied in the solution. The grids are negatively stained with 0.7% (w/v) aqueous uranyl acetate and examined by electron microscopy. The appearance of filaments of approximately 25 nm (microtubules), 8 nm (actin), or 10 nm (intermediate filaments) is a demonstration of MDDT activity.

In another alternative, MDDT activity is measured by the binding of MDDT to protein filaments. ^{35}S-Met labeled MDDT sample is incubated with the appropriate filament protein (actin, tubulin, or intermediate filament protein) and complexed protein is collected by immunoprecipitation using an antibody against the filament protein. The immunoprecipitate is then run out on SDS-PAGE and the amount of MDDT bound is measured by autoradiography.

MDDT activity is measured by its ability to stimulate transcription of a reporter gene (Liu, H.Y. et al. (1997) EMBO J. 16:5289-5298). The assay entails the use of a well characterized reporter gene construct, LexA$_{op}$-LacZ, that consists of LexA DNA transcriptional control elements (LexA$_{op}$) fused to sequences encoding the E. coli LacZ enzyme. The methods for constructing and expressing fusion genes, introducing them into cells, and measuring LacZ enzyme activity, are well known to those skilled in the art. Sequences encoding MDDT are cloned into a plasmid that directs the synthesis of a fusion protein, LexA-MDDT, consisting of MDDT and a DNA binding domain

derived from the LexA transcription factor. The resulting plasmid, encoding a LexA-MDDT fusion protein, is introduced into yeast cells along with a plasmid containing the $LexA_{op}$-LacZ reporter gene. The amount of LacZ enzyme activity associated with LexA-MDDT transfected cells, relative to control cells, is proportional to the amount of transcription stimulated by the MDDT.

Alternatively, MDDT activity is measured by its ability to bind zinc. A 5-10 mM sample solution in 2.5 mM ammonium acetate solution at pH 7.4 is combined with 0.05 M zinc sulfate solution (Aldrich, Milwaukee WI) in the presence of 100 mM dithiothreitol with 10% methanol added. The sample and zinc sulfate solutions are allowed to incubate for 20 minutes. The reaction solution is passed through a VYDAC column (Grace Vydac, Hesperia, CA) with approximately 300 Angstrom bore size and 5 mM particle size to isolate zinc-sample complex from the solution, and into a mass spectrometer (PE Sciex, Ontario, Canada). Zinc bound to sample is quantified using the functional atomic mass of 63.5 Da observed by Whittal, R. M. et al. ((2000) Biochemistry 39:8406-8417).

In the alternative, a method to determine nucleic acid binding activity of MDDT involves a polyacrylamide gel mobility-shift assay. In preparation for this assay, MDDT is expressed by transforming a mammalian cell line such as COS7, HeLa or CHO with a eukaryotic expression vector containing MDDT cDNA. The cells are incubated for 48-72 hours after transformation under conditions appropriate for the cell line to allow expression and accumulation of MDDT. Extracts containing solubilized proteins can be prepared from cells expressing MDDT by methods well known in the art. Portions of the extract containing MDDT are added to $[^{32}P]$-labeled RNA or DNA. Radioactive nucleic acid can be synthesized in vitro by techniques well known in the art. The mixtures are incubated at 25°C in the presence of RNase- and DNase-inhibitors under buffered conditions for 5-10 minutes. After incubation, the samples are analyzed by polyacrylamide gel electrophoresis followed by autoradiography. The presence of a band on the autoradiogram indicates the formation of a complex between MDDT and the radioactive transcript. A band of similar mobility will not be present in samples prepared using control extracts prepared from untransformed cells.

In the alternative, a method to determine methylase activity of MDDT measures transfer of radiolabeled methyl groups between a donor substrate and an acceptor substrate. Reaction mixtures (50 μl final volume) contain 15 mM HEPES, pH 7.9, 1.5 mM $MgCl_2$, 10 mM dithiothreitol, 3% polyvinylalcohol, 1.5 μCi [*methyl*-^{3}H]AdoMet (0.375 μM AdoMet) (DuPont-NEN), 0.6 μg MDDT, and acceptor substrate (e.g., 0.4 μg [^{35}S]RNA, or 6-mercaptopurine (6-MP) to 1 mM final concentration). Reaction mixtures are incubated at 30°C for 30 minutes, then 65°C for 5 minutes.

Analysis of [*methyl*-^{3}H]RNA is as follows: (1) 50 μl of 2 x loading buffer (20 mM Tris-HCl, pH 7.6, 1 M LiCl, 1 mM EDTA, 1% sodium dodecyl sulphate (SDS)) and 50 μl oligo d(T)-cellulose

(10 mg/ml in 1 x loading buffer) are added to the reaction mixture, and incubated at ambient temperature with shaking for 30 minutes. (2) Reaction mixtures are transferred to a 96-well filtration plate attached to a vacuum apparatus. (3) Each sample is washed sequentially with three 2.4 ml aliquots of 1 x oligo d(T) loading buffer containing 0.5% SDS, 0.1% SDS, or no SDS. (4) RNA is eluted with 300 μl of water into a 96-well collection plate, transferred to scintillation vials containing liquid scintillant, and radioactivity determined.

Analysis of [*methyl*-^{3}H]6-MP is as follows: (1) 500 μl 0.5 M borate buffer, pH 10.0, and then 2.5 ml of 20% (v/v) isoamyl alcohol in toluene are added to the reaction mixtures. (2) The samples are mixed by vigorous vortexing for ten seconds. (3) After centrifugation at 700g for 10 minutes, 1.5 ml of the organic phase is transferred to scintillation vials containing 0.5 ml absolute ethanol and liquid scintillant, and radioactivity determined. (4) Results are corrected for the extraction of 6-MP into the organic phase (approximately 41%).

In the alternative, type I topoisomerase activity of MDDT can be assayed based on the relaxation of a supercoiled DNA substrate. MDDT is incubated with its substrate in a buffer lacking Mg^{2+} and ATP, the reaction is terminated, and the products are loaded on an agarose gel. Altered topoisomers can be distinguished from supercoiled substrate electrophoretically. This assay is specific for type I topoisomerase activity because Mg^{2+} and ATP are necessary cofactors for type II topoisomerases.

Type II topoisomerase activity of MDDT can be assayed based on the decatenation of a kinetoplast DNA (KDNA) substrate. MDDT is incubated with KDNA, the reaction is terminated, and the products are loaded on an agarose gel. Monomeric circular KDNA can be distinguished from catenated KDNA electrophoretically. Kits for measuring type I and type II topoisomerase activities are available commercially from Topogen (Columbus OH).

ATP-dependent RNA helicase unwinding activity of MDDT can be measured by the method described by Zhang and Grosse (1994; Biochemistry 33:3906-3912). The substrate for RNA unwinding consists of ^{32}P-labeled RNA composed of two RNA strands of 194 and 130 nucleotides in length containing a duplex region of 17 base-pairs. The RNA substrate is incubated together with ATP, Mg^{2+}, and varying amounts of MDDT in a Tris-HCl buffer, pH 7.5, at 37°C for 30 minutes. The single-stranded RNA product is then separated from the double-stranded RNA substrate by electrophoresis through a 10% SDS-polyacrylamide gel, and quantitated by autoradiography. The amount of single-stranded RNA recovered is proportional to the amount of MDDT in the preparation.

In the alternative, MDDT function is assessed by expressing the sequences encoding MDDT at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a

mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include pCMV SPORT (Life Technologies) and pCR3.1 (Invitrogen Corporation, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, preferably of endothelial or hematopoietic origin, using either liposome formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected.

Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; CLONTECH), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties.

FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M. G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of MDDT on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding MDDT and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Inc., Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding MDDT and other genes of interest can be analyzed by northern analysis or microarray techniques.

Pseudouridine synthase activity of MDDT is assayed using a tritium (^{3}H) release assay modified from Nurse et al. ((1995) RNA 1:102-112), which measures the release of ^{3}H from the C_5 position of the pyrimidine component of uridylate (U) when ^{3}H-radiolabeled U in RNA is isomerized to pseudouridine (y). A typical 500 μl assay mixture contains 50 mM HEPES buffer (pH 7.5), 100 mM

ammonium acetate, 5 mM dithiothreitol, 1 mM EDTA, 30 units RNase inhibitor, and 0.1-4.2 μM [5-^{3}H]tRNA (approximately 1 μCi/nmol tRNA). The reaction is initiated by the addition of <5 μl of a concentrated solution of MDDT (or sample containing MDDT) and incubated for 5 min at 37 °C. Portions of the reaction mixture are removed at various times (up to 30 min) following the addition of MDDT and quenched by dilution into 1 ml 0.1 M HCl containing Norit-SA3 (12% w/v). The quenched reaction mixtures are centrifuged for 5 min at maximum speed in a microcentrifuge, and the supernatants are filtered through a plug of glass wool. The pellet is washed twice by resuspension in 1 ml 0.1 M HCl, followed by centrifugation. The supernatants from the washes are separately passed through the glass wool plug and combined with the original filtrate. A portion of the combined filtrate is mixed with scintillation fluid (up to 10 ml) and counted using a scintillation counter. The amount of ^{3}H released from the RNA and present in the soluble filtrate is proportional to the amount of peudouridine synthase activity in the sample (Ramamurthy, V. (1999) J. Biol. Chem. 274:22225-22230).

In the alternative, pseudouridine synthase activity of MDDT is assayed at 30 °C to 37 °C in a mixture containing 100 mM Tris-HCl (pH 8.0), 100 mM ammonium acetate, 5 mM $MgCl_2$, 2 mM dithiothreitol, 0.1 mM EDTA, and 1-2 fmol of [^{32}P]-radiolabeled runoff transcripts (generated <u>in vitro</u> by an appropriate RNA polymerase, i.e., T7 or SP6) as substrates. MDDT is added to initiate the reaction or omitted from the reaction in control samples. Following incubation, the RNA is extracted with phenol-chloroform, precipitated in ethanol, and hydrolyzed completely to 3-nucleotide monophosphates using RNase T_2. The hydrolysates are analyzed by two-dimensional thin layer chromatography, and the amount of ^{32}P radiolabel present in the yMP and UMP spots are evaluated after exposing the thin layer chromatography plates to film or a PhosphorImager screen. Taking into account the relative number of uridylate residues in the substrate RNA, the relative amount yMP and UMP are determined and used to calculate the relative amount of y per tRNA molecule (expressed in mol y /mol of tRNA or mol y /mol of tRNA/minute), which corresponds to the amount of pseudouridine synthase activity in the MDDT sample (Lecointe, F. et al. (1998) J. Biol. Chem. 273:1316-1323).

N^2,N^2-dimethylguanosine transferase (($m^2{}_2G$)methyltransferase) activity of MDDT is measured in a 160 μl reaction mixture containing 100 mM Tris-HCl (pH 7.5), 0.1 mM EDTA, 10 mM $MgCl_2$, 20 mM NH_4Cl, 1mM dithiothreitol, 6.2 μM *S*-adenosyl-L-[*methyl*-^{3}H]methionine (30-70 Ci/mM), 8 μg $m^2{}_2G$-deficient tRNA or wild type tRNA from yeast, and approximately 100 μg of purified MDDT or a sample comprising MDDT. The reactions are incubated at 30 °C for 90 min and chilled on ice. A portion of each reaction is diluted to 1 ml in water containing 100 μg BSA. 1 ml of 2

M HCl is added to each sample and the acid insoluble products are allowed to precipitate on ice for 20 min before being collected by filtration through glass fiber filters. The collected material is washed several times with HCl and quantitated using a liquid scintillation counter. The amount of ^3H incorporated into the $m^2{}_2G$-deficient, acid-insoluble tRNAs is proportional to the amount of N^2,N^2-dimethylguanosine transferase activity in the MDDT sample. Reactions comprising no substrate tRNAs, or wild-type tRNAs that have already been modified, serve as control reactions which should not yield acid-insoluble ^3H-labeled products.

Polyadenylation activity of MDDT is measured using an in vitro polyadenylation reaction. The reaction mixture is assembled on ice and comprises 10 μl of 5 mM dithiothreitol, 0.025% (v/v) NONIDET P-40, 50 mM creatine phosphate, 6.5% (w/v) polyvinyl alcohol, 0.5 unit/μl RNAGUARD (Pharmacia), 0.025 μg/μl creatine kinase, 1.25 mM cordycepin 5'-triphosphate, and 3.75 mM $MgCl_2$, in a total volume of 25 μl. 60 fmol of CstF, 50 fmol of CPSF, 240 fmol of PAP, 4 μl of crude or partially purified CF II and various amounts of amounts CF I are then added to the reaction mix. The volume is adjusted to 23.5 μl with a buffer containing 50 mM TrisHCl, pH 7.9, 10% (v/v) glycerol, and 0.1 mM Na-EDTA. The final ammonium sulfate concentration should be below 20 mM. The reaction is initiated (on ice) by the addition of 15 fmol of ^{32}P-labeled pre-mRNA template, along with 2.5 μg of unlabeled tRNA, in 1.5 μl of water. Reactions are then incubated at 30 °C for 75-90 min and stopped by the addition of 75 μl (approximately two-volumes) of proteinase K mix (0.2 M Tris-HCl, pH 7.9, 300 mM NaCl, 25 mM Na-EDTA, 2% (w/v) SDS), 1 μl of 10 mg/ml proteinase K, 0.25 μl of 20 mg/ml glycogen, and 23.75 μl of water). Following incubation, the RNA is precipitated with ethanol and analyzed on a 6% (w/v) polyacrylamide, 8.3 M urea sequencing gel. The dried gel is developed by autoradiography or using a phosphoimager. Cleavage activity is determined by comparing the amount of cleavage product to the amount of pre-mRNA template. The omission of any of the polypeptide components of the reaction and substitution of MDDT is useful for identifying the specific biological function of MDDT in pre-mRNA polyadenylation (Rüegsegger, U. et al. (1996) J. Biol. Chem. 271:6107-6113; and references within).

tRNA synthetase activity is measured as the aminoacylation of a substrate tRNA in the presence of [^{14}C]-labeled amino acid. MDDT is incubated with [^{14}C]-labeled amino acid and the appropriate cognate tRNA (for example, [^{14}C]alanine and $tRNA^{ala}$) in a buffered solution. ^{14}C-labeled product is separated from free [^{14}C]amino acid by chromatography, and the incorporated ^{14}C is quantified by scintillation counter. The amount of ^{14}C-labeled product detected is proportional to the activity of MDDT in this assay.

In the alternative, MDDT activity is measured by incubating a sample containing MDDT in a

solution containing 1 mM ATP, 5 mM Hepes-KOH (pH 7.0), 2.5 mM KCl, 1.5 mM magnesium chloride, and 0.5 mM DTT along with misacylated [^{14}C]-Glu-tRNAGln (e.g., 1 μM) and a similar concentration of unlabeled L-glutamine. Following the quenching of the reaction with 3 M sodium acetate (pH 5.0), the mixture is extracted with an equal volume of water-saturated phenol, and the aqueous and organic phases are separated by centrifugation at 15,000 × g at room temperature for 1 min. The aqueous phase is removed and precipitated with 3 volumes of ethanol at -70°C for 15 min. The precipitated aminoacyl-tRNAs are recovered by centrifugation at 15,000 × g at 4°C for15 min. The pellet is resuspended in of 25 mM KOH, deacylated at 65°C for 10 min., neutralized with 0.1 M HCl (to final pH 6-7), and dried under vacuum. The dried pellet is resuspended in water and spotted onto a cellulose TLC plate. The plate is developed in either isopropanol/formic acid/water or ammonia/water/chloroform/ methanol. The image is subjected to densitometric analysis and the relative amounts of Glu and Gln are calculated based on the Rf values and relative intensities of the spots. MDDT activity is calculated based on the amount of Gln resulting from the transformation of Glu while acylated as Glu-tRNA^{Gln} (adapted from Curnow, A.W. et al. (1997) Proc. Natl. Acad. Sci. 94:11819-26).

XIX. Identification of MDDT Agonists and Antagonists

Agonists or antagonists of MDDT activation or inhibition may be tested using the assays described in section XVII. Agonists cause an increase in MDDT activity and antagonists cause a decrease in MDDT activity.

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	CA2 Reagents
71230017	1	71230017CD1	24	71230017CB1	
3125036	2	3125036CD1	25	3125036CB1	
1758089	3	1758089CD1	26	1758089CB1	
3533891	4	3533891CD1	27	3533891CB1	
1510943	5	1510943CD1	28	1510943CB1	
2119377	6	2119377CD1	29	2119377CB1	2119377CA2
3176058	7	3176058CD1	30	3176058CB1	
2299818	8	2299818CD1	31	2299818CB1	90135665CA2
2729451	9	2729451CD1	32	2729451CB1	
878534	10	878534CD1	33	878534CB1	
2806157	11	2806157CD1	34	2806157CB1	2806157CA2, 7976113CA2
5883626	12	5883626CD1	35	5883626CB1	2201431CA2, 2957907CA2, 5890236CA2, 5891113CA2, 5891191CA2
2674016	13	2674016CD1	36	2674016CB1	
5994159	14	5994159CD1	37	5994159CB1	3564793CA2
2457335	15	2457335CD1	38	2457335CB1	
2267802	16	2267802CD1	39	2267802CB1	
3212060	17	3212060CD1	40	3212060CB1	3591224CA2
3121069	18	3121069CD1	41	3121069CB1	3142557CA2
3280626	19	3280626CD1	42	3280626CB1	
484404	20	484404CD1	43	484404CB1	
2830063	21	2830063CD1	44	2830063CB1	
7506096	22	7506096CD1	45	7506096CB1	
7505914	23	7505914CD1	46	7505914CB1	

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
1	71230017CD1	g15982946	0.0	SSA protein SS-56 [Homo sapiens] Billaut-Mulot,O. et al. (2001) SS-56, a novel cellular target of autoantibody responses in Sjogren syndrome and systemic lupus erythematosus. J. Clin. Invest. 108:861-869
2	3125036CD1	g5690435	4.0E-116	[Xenopus laevis] nuclear protein Sojo
		g10432382	1.7E-234	[Homo sapiens] dJ717I23.1 (novel protein similar to Xenopus laevis Sojo protein)
3	1758089CD1	g10567164	0.0	[Homo sapiens] gene amplified in squamous cell carcinoma-1 Yang, Z.Q. et al. (2000) Cancer Res. 60:4735-4739
4	3533891CD1	g5823146	2.9E-74	[Rattus norvegicus] testis specific protein
5	1510943CD1	g13604149	0.0	tangerin C' [Mus musculus]
6	2119377CD1	g18034072	1.0E-122	SPRY domain-containing SOCS box protein SSB-1 [Homo sapiens]
9	2729451CD1	g12856615	1.0E-144	DNA BINDING PROTEIN DESRT~data source:SPTR, source key:Q9JIX4, evidence:ISS~putative [Mus musculus] Carninci,P. and Hayashizaki,Y. (1999) High-efficiency full-length cDNA cloning. Meth. Enzymol. 303:19-44 Carninci,P. et al. (2000) Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10:1617-1630
11	2806157CD1	g2587026	2.7E-32	[Homo sapiens] HERV-E integrase Lindeskog, M. et al. (1998) Virology 244:219-229
14	5994159CD1	g7768636	3.5E-31	[Xenopus laevis] Kielin Matsui, M. et al. (2000) Proc. Natl. Acad. Sci. USA 97:5291-5296
		g6979313	2.0E-16	cysteine-rich repeat-containing protein CRIM1 [Mus musculus]
15	2457335CD1	g12584947	8.3E-134	[Homo sapiens] ovary-specific acidic protein
16	2267802CD1	g12963885	0.0	[Homo sapiens] (AY026527) prostate antigen PARIS-1
21	2830063CD1	g13539684	0.0	zinc finger protein 291 [Homo sapiens]

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
22	7506096CD1	g2773363	1.4E-49	[Drosophila melanogaster] microtubule binding protein D-CLIP-190 Lantz, V.A. and Miller, K.G. (1998) A class VI unconventional myosin is associated with a homologue of a microtubule-binding protein, cytoplasmic linker protein-170, in neurons and at the posterior pole of Drosophila embryos. J. Cell Biol. 140:897-910
		339768|CENPE	1.4E-49	[Homo sapiens][Motor protein; Hydrolase; ATPase][Nuclear] Centromere protein E, a kinesin-like minus-end directed motor protein, associated with kinetochores, required for chromosome alignment during metaphase and metaphase to anaphase transition, may have a role in rheumatoid arthritis and systematic sclerosis. Kullmann, F. et al. (1999) Arthritis Res. 1:71-80
		568434|GOLGA4	1.7E-48	[Homo sapiens][Golgi; Cytoplasmic; Plasma membrane] Golgi autoantigen golgin subfamily a 4 (golgin-245), contains a novel Golgi-targeting GRIP domain, may function in vesicular transport from the trans-Golgi, vesicle biogenesis, or Golgi structural organization; autoantigen in Sjogren's syndrome patients.
		335126|EEA1	4.4E-45	[Homo sapiens][Small molecule-binding protein] [Endosome/Endosomal vesicles; Nuclear; Cytoplasmic; Plasma membrane] Early endosome antigen 1, effector of endosomal small GTPase RAB5, required for endosome fusion, may specify transport directionality from the plasma membrane to early endosomes; autoantigen associated with subacute cutaneous systemic lupus erythematosus. Mu, F. T. et al. (1995) J. Biol. Chem. 270:13503-13511.
23	7505914CD1	g18642530	0.0	SR rich protein [Homo sapiens]

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
23		610045|Srrp86	3.2E-27	[Rattus norvegicus][Spliceosomal subunit; RNA-binding protein] [Nuclear] Serine arginine-rich splicing regulatory protein 86, contains an RNA recognition motif and serine-arginine-rich domains, interacts with other serine-arginine-rich splicing factors, involved in splicing regulation and differential splice site selection Barnard, D. C., and Patton, J. G. (2000) Identification and characterization of a novel serine-arginine-rich splicing regulatory protein. Mol. Cell. Biol. 20:3049-3057

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	71230017CD1	485	S183 S252 S355 T170 T172 T179 Y313	N230 N268 N438 N471	Signal peptide: M1-S50	SPScan
					SPRY domain: S355-D482	HMMER-PFAM
					B-box zinc finger.: L93-M134	HMMER-PFAM
					Zinc finger, C3HC4 type (RING finger): C16-C60	HMMER-PFAM
					Zinc finger, C3HC4 type (RING finger), signature: I10-R67	ProfileScan
					Zinc finger, C3HC4 type: C31-C39	BLIMPS-BLOCKS
					Domain in SPla: PF00622A: K110-S123 PF00622B: E339-W360 PF00622C: V423-F436	BLIMPS-PFAM
					Midline zinc finger, RING, stonus toxin, putative transcription factor PD002421: L298-F462	BLAST-PRODOM
					Butyrophilin, zinc finger, DNA-binding PD002445: L260-Q351	BLAST-PRODOM
					Receptor, ryanodine, transmembrane, calcium channel, butyrophilin PD001178: S355-F449	BLAST-PRODOM
					RFP transforming protein DM02346: P19474|59-337: R67-Q351 A57041|64-348: Q65-G356 P14373|61-366: R67-C352	BLAST-DOMO
					RFP transforming protein DM01944: P19474|339-465: S355-D482	BLAST-DOMO
					Zinc finger, C3HC4 type (RING finger), signature: C31-L40	MOTIFS
					Leucine zipper pattern: L227-L248	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
2	3125036CD1	1404	S4 S24 S38 S47 S59 S61 S79 S90 S115 S156 S183 S199 S209 S213 S316 S365 S407 S408 S444 S500 S504 S521 S587 S588 S599 S680 S711 S727 S771 S783 S831 S852 S927 S1005 S1018 S1096 S1119 S1164 S1169 S1180 S1194 S1256 S1273 S1305 S1336 S1341 S1352 S1391	N134 N296 N481 N495 N586 N725 N1344	Coiled coil protein, myosin repeat, heavy, ATP-binding, filament, heptad PD000002: L878-L1127	BLAST-PRODOM
			T139 T283 T298 T493 T543 T595 T645 T753 T764 T815 T861 T863 T882 T910 T934 T978 T983 T1310 T1337 T1348 Y243 Y715		Tropomyosin repeat, coiled coil PD000023: N870-S1096	BLAST-PRODOM
					Coiled coil, heptad repeat, ATP-binding PD075049: L865-D1123	BLAST-PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
2					Dynein chain, motor, microtubules, ATP-binding, heptad repeat PD003395: H568-D1263	BLAST-PRODOM
					Trichohyalin DM03839|P37709|632-1103: Q739-D1193	BLAST-DOMO
					Heptad repeat pattern: DM05319|P30427|568-1938: K532-L1345	BLAST-DOMO
					Leucine zipper pattern: L116-L137, L900-L921, L907-L928	MOTIFS
3	1758089CD1	1096	S12 S104 S140 S153 S364 S373 S378 S407 S452 S458 S483 S566 S610 S632 S633 S641 S647 S707 S735 S863 S956 S978 S1051 S1072	N125	PHD-finger: G750-H791, K851-Y897	HMMER-PFAM
			T17 T21 T59 T94 T109 T156 T167 T294 T308 T340 T351 T560 T571 T699 T811 T946 T967 T1017 T1025 Y993		jmjC domain: Y176-F292	HMMER-PFAM
					jmjN domain: K14-D61	HMMER-PFAM
					PHD-finger: Y871-A885	BLIMPS-PFAM
					XE169, nuclear, zinc finger, DNA-binding PROTEIN INTERGENIC REGION XE169 PD005470: E97-R329	BLAST-PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
3					zinc finger, nuclear, DNA-binding, ALL1, translocation, protooncogene PD006688: E796-H906	BLAST-PRODOM
					Finger, SMCX, SMCY, YDR096W, DM01930: P39956|83-380: L118-M318 P29375|346-638: W149-C307 S44139|245-535: W149-C307 P41229|377-669: W149-C307	BLAST-DOMO
					Cell attachment sequence RGD: R1020-D1022	MOTIFS
4	3533891CD1	167	S64 S70 S89 S122 S163 T8 T101	N42	Signal peptide: M38-A91	SPScan
5	1510943CD1	1523	S141 S176 S191 S239 S264 S290 S310 S337 S361 S390 S533 S714 S852 S993 S998 S1016 S1042 T32 S1065 S1123 T45 S1168 S1257 S1288 S1297 S1338 S1346 S1390 S1511 S1515	N104 N967 N1061 N1292	Calponin homology (CH) domain: V1037-T1142	HMMER-PFAM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
5			T164 T219 T258 T284 T405 T470 T521 T572 T646 T653 T669 T704 T730 T866 T971 T1142 T1159 T1326 Y1086 Y1362		Transmembrane domain: E755-G771 N-terminus is non-cytosolic	TMAP
					Alpha-actinin actin-binding domain DM00325: P18091|28-252: V1037-F1140 Q08043|39-263: S1038-F1140 A44159|48-277: S1038-L1134 P35609|32-256: S1038-F1140	BLAST-DOMO
					Leucine zipper pattern: L1404-L1425	MOTIFS
					Binding-protein-dependent transport systems inner membrane comp. signal: V1207-P1235	MOTIFS
6	2119377CD1	273	S8 S135 S244 S265 T119 T223		Signal peptide: M1-A55	SPScan
					Mouse BAC library, BAC284H12 12P13 PD039422: P34-Q273	BLAST-PRODOM
					Trp-Asp (WD) repeats signature: T130-L144	MOTIFS
7	3176058CD1	341	S10 S80 S136 S191 S204 S218 S269 T155 T196	N75 N153	C11D2.4 protein PD137800: M1-R337	BLAST-PRODOM
8	2299818CD1	341	S45 S78 S91 Y97 S169 S203 S328 T33 T192 T281	N7 N31 N201 N263 N331 N336	Signal peptide: M1-D37	SPScan
					Protein HES1SEC63 B0024.11 409AA PD005058: K4-L186, R130-I258	BLAST-PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
9	2729451CD1	1185	S37 S72 S99 S239 S264 S304 S428 S451 S469 S480 S483 S504 S510 S524 S526 S573 S715 S754 S776 S869 S972 S999 S1012 S1029 S1038 S1044 S1150 S1182	N237 N273 N427 N434 N518 N606 N622 N864 N1105	ARID DNA binding domain: E315-E426	HMMER-PFAM
			T47 T65 T233 T337 T369 T394 T441 T608 T624 T642 T765 T850 T915 Y344		Transmembrane domain: I201-V216 N-terminus is non-cytosolic	TMAP
					Nuclear DNA-binding protein, transcription, DRIL1, retinoblastoma, trans-acting factor PD004601: F324-P416	BLAST-PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
10	878534CD1	1042	S95 S168 S245 S276 S337 S375 S407 S411 S434 S457 S535 S565 S582 S598 S614 S659 S704 S714 S718 S795 S826 S834 S838 S882 S884 S916 S925 S958 S1005 T49 T68 T162 T166 T347 T362 T419 T508 T622 T765 T811 T812 T946 T1001 T1009 T1040	N47 N142 N172 N207 N225 N226 N230 N620	Signal peptide: M1-A34	SPScan
11	2806157CD1	86	T72 T77 T83		Similar to HERV H protease and HERV E integrase protease PD064787: P53-S86	BLAST-PRODOM
12	5883626CD1	138	S24 S68		Signal peptide: M4-D71	SPScan
					Transmembrane domain: C53-C69 N-terminus is non-cytosolic	TMAP

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
13	2674016CD1	805	S30 S52 S68 S204 S264 S286 S290 S305 S321 S396 S401 S408 S467 S491 S542 S546 S551 S559 S577 S584 S597 S619 S653 S705 S706 S717 S728 S736 S740 S748 S752 S757 S760 S767 S787	N487 N648	DNA-binding protein PD001830: K581-K799, K553-S783, R594-S804	BLAST-PRODOM
			S790 S795 T231 T271 T326 T350 T366 T410 T448 T485 T565 T628 T744		Topoisomerase I, DNA isomerase, DNA-binding, intermediate filament heptad PD000422: E603-R796, R640-K797	BLAST-PRODOM
					Type B repeat DM05511: S26650|1-1203: E462-T745, K500-R803, R472-S760 P18583|113-1296: E462-T745, G506-R803, D402-K675	BLAST-DOMO
					Caldesmon: DM06224|P12957|1-755: S405-S779, A193-K750	BLAST-DOMO
					Tumor recognition, prolyl: DM08077|P30414|230-1403: E481-S804, E603-S802, E244-E324	BLAST-DOMO
14	5994159CD1	426	S72 S115 S133 S212 S218 S312 S373 S419 T103 T172 T396	N110 N250	Signal peptide: M22-S72	SPScan

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
14					von Willebrand factor type C domain: C158-C213, C100-C155	HMMER-PFAM
					von Willebrand factor type C repeat DM00551|A38963|649-756: R59-C155	BLAST-DOMO
					C-type lectin domain: C120-C141	MOTIFS
					von Willebrand factor C domain signature: C120-C155, C178-C213	MOTIFS
15	2457335CD1	267	S29 S34 S35 S213 S220 T85 T102 T156 T175 T196 T197 T201	N199	Signal peptide: M1-A58	SPScan
					Transmembrane domain: N36-Y64	TMAP
16	2267802CD1	928	S21 S207 S253 S267 S324 S346 S391 S422 S558 S690 S756 S768 S859 S909 S920 T83 T121 T231 T303 T567 Y815	N205 N288 N301 N675	PH domain: K46-W142	HMMER-PFAM
					TBC domain: L622-L839	HMMER-PFAM
					Transmembrane domain: V783-L806 N-terminus is cytosolic	TMAP
					Probable rabGAP domain PF00566: I670-P679, Y711-N716	BLIMPS-PFAM
					Transmembrane protein, cell division, oncogene PD001799: D693-L843	BLAST-PRODOM

Table 3

<table>
<tr><th>SEQ ID NO:</th><th>Incyte Polypeptide ID</th><th>Amino Acid Residues</th><th>Potential Phosphorylation Sites</th><th>Potential Glycosylation Sites</th><th>Signature Sequences, Domains and Motifs</th><th>Analytical Methods and Databases</th></tr>
<tr><td>16</td><td></td><td></td><td></td><td></td><td>Membrane protein DM01737 S62481|395-698: E617-R823 Q09830|395-698: E617-R823 P53258|152-437: L612-R823 P48566|107-461: R533-H782, L785-R823</td><td>BLAST-DOMO</td></tr>
<tr><td>17</td><td>3212060CD1</td><td>684</td><td>S116 S121 S194 S232 S254 S369 S382 S419 S493 S576 S653 S654 S680 T44 T56 T189 T263 T496 T529 T679 Y233 Y484</td><td>N273 N351</td><td>DnaB-like helicase PF00772: L390-Y428, T439-Y471, I510-M521, T56-K91</td><td>BLIMPS-PFAM</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Similarity to ATP/GTP-binding site motif A PD145092: E153-A460, W474-S629</td><td>BLAST-PRODOM</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Cell attachment sequence: R132-D134</td><td>MOTIFS</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>ATP/GTP-binding site motif A (P-loop): G415-T422</td><td>MOTIFS</td></tr>
<tr><td>18</td><td>3121069CD1</td><td>267</td><td>S36 S98 T27 T86 S183 S219 S232 S234 T137 T141 T156 T203 T249</td><td>N45 N54 N82 N114 N128 N135 N154 N179</td><td>Transmembrane domains: T4-T27, T181-I207 N-terminus is cytosolic</td><td>TMAP</td></tr>
<tr><td>19</td><td>3280626CD1</td><td>537</td><td>S37 S123 S137 S267 S274 S308 S314 S438 S456 T157 T171 T320 T477 Y190 Y246 Y483</td><td>N312 N318</td><td></td><td></td></tr>
</table>

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
20	484404CD1	312	S55 S85 S95 S117 S123 S142 S159 S198 S304 T32 T145 T170 T233 Y250	N219		
21	2830063CD1	1400	S37 S52 S126 S177 S221 S254 S294 S298 S349 S359 S417 S502 S508 S535 S543 S753 S773 S832 S840 S888 S895 S930 S1011 S1090 S1159 S1203 S1219 S1290	N192 N468 N506 N823 N995 N1000 N1004 N1033 N1087 N1207	Transmembrane domains: N1040-R1068, I1103-L1120, A1133-V1153, S1159-D1179, H1185-K1205, Q1214-S1236 N-terminus is non-cytosolic	TMAP
			T24 T59 T62 T67 T77 T188 T350 T466 T539 T566 T786 T935 T961 T1041 T1077 T1154 T1195 Y118 Y1105		Coiled coil, myosin repeat, ATP-binding, heptad PD000002: M527-K767, E529-K749, Q570-E770	BLAST-PRODOM
					Coiled coil, tropomyosin repeat PD000023: K568-E770, R536-Q763	BLAST-PRODOM
					Trichohyalin DM03839: P37709|632-1103: A400-K767, E542-N894, R536-D920 Q07283|91-443: E501-L771, V493-K767 P22793|921-1475: R538-K767, E529-N947	BLAST-DOMO BLAST-DOMO BLAST-DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
21					Tropomyosin DM00077|P37709|1104-1277: K545-R727, K545-Q705, E596-K767	BLAST-DOMO
					Zinc finger, C2H2 type, domain: C794-H816	MOTIFS
22	7506096CD1	1384	S4 S24 S38 S47 S59 S61 S79 S90 S115 S156 S183 S199 S209 S213 S296 S345 S387 S388 S424 S480 S484 S501 S567 S568 S579 S660 S691 S707 S751 S763 S811 S832 S907 S985 S998 S1076 S1099 S1144 S1149 S1160 S1174 S1236 S1253 S1285 S1316 S1321 S1332 S1371	N134 N276 N461 N475 N566 N705 N1324	PROTEIN COILED COIL CHAIN MYOSIN REPEAT HEAVY ATP-BINDING FILAMENT HEPTAD PD000002: L858-L1107, K569-K815, Q133-K357	BLAST_PRODOM
			T139 T278 T473 T523 T575 T625 T733 T744 T795 T841 T843 T862 T890 T914 T958 T963 T1290 T1317 T1328 Y243 Y695		MYOSIN MYOSIN 3 ISOFORM HEAVY CHAIN TYPE II COILED COIL ATP-BINDING PD031043: L242-E1212	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
22					PROTEIN REPEAT TROPOMYOSIN COILED COIL ALTERNATIVE SPLICING SIGNAL PRECURSOR CHAIN PD000023: N850-S1076	BLAST_PRODOM
					SCABROUS PROTEIN PRECURSOR DEVELOPMENTAL NEUROGENESIS SIGNAL PD144674: V182-K629	BLAST_PRODOM
					MYOSIN-LIKE PROTEIN MLP1 DM07884|Q02455|35-1728: M1-L1325	BLAST_DOMO
					TRICHOHYALIN DM03839|P37709|632-1103: Q719-D1173, Q185-L658	BLAST_DOMO
					HEPTAD REPEAT PATTERN REPEAT DM05319|P30427|568-1938: L210-I1235	BLAST_DOMO
					Leucine zipper pattern: L116-L137, L880-L901, L887-L908	MOTIFS
23	7505914CD1	787	S30 S52 S68 S204 S264 S286 S290 S305 S321 S378 S383 S390 S449 S473 S524 S528 S533 S541 S559 S566 S579 S601 S635 S687 S688 S699 S710 S718 S722 S730 S734 S739 S742 S749 S769 S772	N469 N630	signal_cleavage: M1-A47	SPSCAN

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
23			S777 T231 T271 T326 T350 T366 T392 T430 T467 T547 T610 T726		Protamine P1 proteins BL00048: R596-R622	BLIMPS_BLOCKS
					PROTEIN DNA BINDING CODED FOR BY C ELEGANS cDNA CHROMOSOME HOMOLOG PD001830: K563-K781, K535-S765, R576-S784, E485-G735, D459-D683	BLAST_PRODOM
					PROTEIN TOPOISOMERASE I DNA ISOMERASE REPEAT DNA BINDING INTERMEDIATE FILAMENT HEPTAD PD000422: E585-R778, R622-R785	BLAST_PRODOM
					TYPE B REPEAT REPEAT DM05511 |S26650|1-1203: E444-T727, K482-R785, R539-R787 |P18583|113-1296: E444-T727, G488-R785, D384-K657, R539-R787	BLAST_DOMO

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
24/71230017CB1/ 3332	1-236, 1-266, 16-634, 230-686, 243-825, 259-932, 487-759, 513-969, 677-939, 677-1275, 677-1294, 847-1117, 908-1017, 938-1287, 974-1510, 1018-1071, 1018-1255, 1028-1572, 1031-1265, 1040-1677, 1052-1672, 1168-1716, 1218-1836, 1286-1803, 1312-1757, 1354-1734, 1354-1954, 1491-2101, 1494-1950, 1496-2141, 1514-1754, 1514-2053, 1533-1733, 1533-1817, 1555-1848, 1591-1876, 1651-2255, 1651-2296, 1673-2296, 1736-2279, 1741-1962, 1746-2323, 1756-2247, 1757-2361, 1768-2383, 1783-1980, 1796-2429, 1816-2020, 1843-2163, 1845-2482, 1850-2515, 1916-2500, 1923-2581, 1963-2252, 1971-2616, 1978-2161, 1995-2516, 2009-2648, 2013-2679, 2020-2185, 2023-2430, 2034-2672, 2069-2453, 2072-2229, 2079-2341, 2079-2586, 2084-2610, 2092-2367, 2092-2379, 2100-2680, 2106-2556, 2111-2712, 2122-2672, 2147-2769, 2163-2808, 2167-2731, 2187-2737, 2193-2438, 2199-2410, 2199-2414, 2229-2756, 2253-2864, 2271-2564, 2286-2850, 2286-2861, 2391-2951, 2438-3228, 2439-3228, 2495-3228, 2497-3249, 2620-3012, 2624-3193, 2627-2873, 2628-2931, 2629-3244, 2632-2895, 2650-2896,
	2651-2928, 2651-3143, 2652-3130, 2675-3267, 2677-3256, 2682-3231, 2683-3256, 2689-3266, 2692-3237, 2708-3080, 2734-2977, 2734-3067, 2744-3013, 2747-3193, 2776-3085, 2793-3090, 2803-3002, 2804-3063, 2817-3273, 2818-3085, 2837-3139, 2850-3276, 2851-3273, 2854-3065, 2854-3273, 2861-3101, 2871-3280, 2874-3101, 2876-3167, 2878-3276, 2880-3108, 2908-3284, 2919-3270, 2935-3277, 2958-3276, 2960-3275, 2995-3248, 3031-3332, 3036-3273, 3039-3276, 3047-3273, 3097-3322
25/3125036CB1/ 4410	1-134, 1-3732, 285-772, 285-774, 285-779, 288-583, 302-578, 425-797, 525-797, 706-857, 852-901, 854-1003, 1395-3393, 1487-2022, 2023-2309, 2023-3501, 2029-2146, 2138-2386, 2138-2628, 2184-2445, 2185-2735, 2185-2767, 2449-2614, 2469-2688, 2469-2941, 2614-2730, 2651-3318, 2674-3318, 2695-3001, 2695-3096, 2695-3151, 2726-3214, 2737-3341, 2740-3328, 2794-3333, 2799-3341, 2826-2931, 2868-3418, 2899-3408, 2927-3188, 2927-3362, 2927-3364, 2927-3404, 2927-3423, 2986-3141, 2986-3558, 2988-3512, 3006-3226, 3006-3499, 3020-3695, 3043-3250, 3063-3334, 3075-3612, 3100-3457, 3127-3377, 3127-3379, 3127-3393, 3151-3413, 3179-3732, 3181-3705, 3190-3368, 3195-3492, 3216-3404, 3236-3733, 3240-3526, 3242-3922, 3247-3732, 3254-3724, 3339-3732, 3341-3928, 3362-3732, 3436-3732, 3447-3539, 3449-4037, 3530-3603, 3530-3604, 3530-3605, 3530-3710, 3530-3732, 3530-3733, 3531-3603, 3565-3732, 3733-4201, 3802-4376, 3879-4410, 3884-4410, 3924-4084, 3939-4404, 3946-4410, 3991-4409, 3991-4410, 4043-4409, 4169-4344, 4189-4410, 4241-4344

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
26/1758089CB1/ 5032	1-625, 13-519, 19-673, 45-539, 45-540, 117-434, 157-760, 157-882, 161-383, 176-896, 227-839, 316-931, 401-1057, 401-1061, 401-1065, 401-1098, 430-467, 539-718, 555-1108, 603-1147, 611-1311, 717-1311, 797-1214, 921-1535, 929-1558, 961-1114, 1028-5022, 1049-1311, 1207-1861, 1312-1360, 1312-1490, 1312-1861, 1334-2183, 1351-1996, 1391-1623, 1459-1966, 1537-2141, 1546-1805, 1673-2193, 1726-2275, 1745-2183, 1799-2499, 1857-2370, 1910-2509, 1916-2509, 2011-2254, 2024-2568, 2097-2748, 2112-2632, 2398-2746, 2399-3091, 2421-3010, 2746-2934, 2798-2934, 2839-3320, 2911-3163, 2923-3023, 2933-3097, 2983-3236, 3001-3634, 3102-3574, 3102-3642, 3120-3489, 3123-3715, 3227-3487, 3297-3871, 3395-3690, 3403-3659, 3403-3857, 3432-3857, 3461-3855, 3466-3859, 3475-3857, 3502-3861, 3533-3763, 3533-3857, 3557-3857, 3596-4142, 3616-4212, 3621-3863, 3648-3895, 3666-4096, 3666-4302, 3668-4015, 3668-4121, 3668-4136, 3673-3826, 3708-3857, 3709-4121, 3732-4060, 3739-4089, 3739-4294, 3744-3967, 3751-4035, 3763-4069, 3764-4061, 3784-3915, 3811-4202, 3831-4068, 3851-4361, 3853-4054, 3853-4116, 3853-4179, 3860-4117, 3870-4149, 3872-4155, 3881-4141, 3881-4457, 3888-4491,
	3922-4218, 3932-4226, 3950-4131, 3950-4229, 3969-4146, 3971-4264, 3989-4257, 3989-4506, 3992-4259, 3993-4262, 3994-4249, 3999-4209, 4002-4239, 4015-4235, 4015-4565, 4016-4276, 4026-4338, 4028-4251, 4035-4245, 4035-4610, 4049-4241, 4052-4326, 4053-4547, 4066-4756, 4104-4191, 4105-4372, 4145-4343, 4184-4426, 4187-4820, 4211-4832, 4224-4496, 4234-4476, 4240-4536, 4246-4523, 4270-4526, 4270-4774, 4273-4610, 4284-5021, 4292-4560, 4304-4774, 4324-4530, 4324-5022, 4357-4618, 4384-4701, 4389-4653, 4393-4907, 4405-4962, 4419-5032, 4424-5032, 4425-4680, 4426-4687, 4462-5014, 4462-5032, 4469-4732, 4514-4810, 4526-5032, 4535-5032, 4562-4780, 4562-4807, 4563-4794, 4571-5032, 4574-4807, 4594-5032, 4608-4807, 4617-4808, 4662-4929, 4665-4938, 4665-5032, 4666-5032, 4678-4807, 4703-5032, 4708-4807, 4730-4805
27/3533891CB1/ 1355	1-660, 3-529, 3-591, 21-664, 51-302, 51-516, 51-621, 382-696, 499-624, 499-1008, 664-934, 705-912, 722-1300, 723-1038, 812-1355

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
28/1510943CB1/ 4912	1-550, 129-711, 153-792, 165-845, 180-845, 205-718, 233-944, 272-689, 300-936, 315-989, 389-828, 407-845, 416-1050, 455-654, 494-4912, 497-1030, 498-988, 646-779, 683-1065, 755-1322, 782-1028, 784-1401, 826-990, 871-1038, 890-1038, 899-1409, 914-990, 980-1038, 989-1229, 1012-1668, 1035-1118, 1035-1140, 1035-1143, 1035-1149, 1055-1140, 1066-1140, 1069-1243, 1073-1711, 1117-1140, 1166-1709, 1213-1660, 1249-1931, 1292-1714, 1295-1922, 1312-1896, 1333-1781, 1337-1834, 1374-1779, 1413-1672, 1452-2172, 1489-1945, 1503-1954, 1521-1943, 1535-1571, 1535-1589, 1535-1593, 1535-1625, 1535-1680, 1535-1683, 1535-1762, 1535-1846, 1535-1935, 1535-1956, 1535-1976, 1535-1996, 1535-2015, 1539-2086, 1543-2043, 1550-2217, 1559-1948, 1563-2186, 1605-1716, 1618-1915, 1618-2011, 1652-2186, 1685-2186, 1687-2186, 1691-1903, 1691-2186, 1697-2186, 1704-2186, 1714-2186, 1732-2034, 1742-2008, 1747-2305, 1750-2186, 1762-2186, 1772-2297, 1776-2186, 1780-2326, 1796-2186, 1802-2368, 1805-2290, 1805-2325, 1808-1842, 1817-2270, 1827-2186, 1848-2325, 1849-1897, 1875-2380,
	1902-2326, 1941-2521, 2005-2328, 2024-2271, 2055-2574, 2090-2832, 2093-2769, 2099-2338, 2106-2132, 2107-2132, 2107-2352, 2137-2158, 2179-2805, 2182-2476, 2190-2376, 2219-2786, 2225-2509, 2240-2725, 2256-2287, 2268-2637, 2272-2378, 2313-2870, 2326-2401, 2337-2378, 2337-2392, 2345-2375, 2371-2791, 2456-3016, 2456-3033, 2460-2502, 2460-2545, 2460-2582, 2460-2583, 2465-2545, 2465-2583, 2468-2560, 2476-2542, 2486-2583, 2487-2526, 2487-2528, 2487-2534, 2487-2542, 2492-2583, 2495-3160, 2519-2583, 2532-2591, 2551-2850, 2595-2644, 2595-2669, 2595-2680, 2595-2718, 2598-2757, 2598-3270, 2603-2675, 2611-2669, 2611-2686, 2613-3222, 2627-2879, 2627-2884, 2627-3159, 2630-2718, 2653-2925, 2653-3186, 2668-3144, 2671-2876, 2671-2881, 2671-2885, 2671-2889, 2787-3007, 2810-3146, 2831-3501, 2836-3444, 2857-3456, 2867-3384, 3004-3244, 3006-3427, 3010-3264, 3012-3281, 3033-3227, 3033-3365, 3051-3336, 3053-3738, 3057-3333, 3062-3121, 3062-3264, 3088-3590, 3099-3701, 3127-3363, 3130-3668, 3138-3752, 3147-3264, 3259-3296, 3261-3779, 3282-3874, 3409-3955, 3409-3995,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
28	3409-3997, 3409-4020, 3591-3666, 3640-4104, 3641-3806, 3693-4116, 3766-4041, 3921-4168, 3921-4303, 3970-4453, 3983-4282, 3984-4096, 3984-4302, 3985-4453, 3994-4264, 3995-4176, 4003-4273, 4003-4304, 4005-4453, 4020-4253, 4020-4273, 4036-4249, 4036-4279, 4037-4272, 4037-4509, 4072-4317, 4081-4645, 4089-4309, 4098-4453, 4106-4430, 4112-4306, 4112-4320, 4112-4323, 4116-4310, 4135-4425, 4140-4411, 4150-4370, 4154-4762, 4166-4430, 4187-4444, 4189-4912, 4214-4912, 4244-4420, 4244-4527, 4251-4912, 4254-4453, 4254-4912, 4268-4527, 4271-4453, 4284-4908, 4305-4824, 4309-4453, 4319-4902, 4328-4453, 4347-4912, 4361-4466, 4363-4661, 4367-4522, 4374-4912, 4377-4657, 4420-4912, 4426-4912, 4427-4912, 4429-4912, 4430-4887, 4430-4895, 4433-4912, 4434-4902, 4437-4896, 4437-4902, 4454-4572, 4454-4587, 4454-4600, 4454-4661, 4454-4695, 4454-4796, 4454-4835, 4454-4853, 4454-4859, 4454-4872, 4454-4873, 4454-4877, 4454-4882, 4454-4883, 4454-4887, 4454-4897, 4454-4901, 4454-4902, 4454-4912, 4455-4885, 4461-4912, 4464-4897, 4466-4702, 4467-4871, 4471-4577, 4471-4902, 4471-4912, 4473-4912,
	4474-4912, 4475-4912, 4476-4912, 4477-4770, 4482-4912, 4483-4912, 4486-4912, 4491-4912, 4493-4912, 4499-4902, 4692-4731, 4788-4912, 4865-4890
29/2119377CB1/ 2241	1-269, 1-461, 1-508, 1-515, 1-554, 1-584, 1-655, 1-707, 1-708, 1-711, 1-716, 2-461, 4-577, 43-663, 69-715, 275-832, 275-913, 312-1048, 331-952, 342-865, 342-915, 343-980, 364-916, 413-947, 423-935, 427-874, 450-1053, 482-1144, 501-962, 501-1210, 552-1193, 552-1360, 560-1218, 562-1218, 567-1218, 567-1264, 591-1210, 592-1245, 593-1221, 594-1332, 604-1249, 613-1116, 638-1317, 639-1246, 639-1254, 648-1308, 670-1034, 680-1374, 701-1375, 704-1496, 711-1431, 717-1393, 744-1603, 853-1383, 901-1444, 921-1256, 924-1435, 975-1557, 1115-1388, 1115-1620, 1141-1754, 1370-1799, 1399-1680, 1440-2027, 1516-1990, 1560-2198, 1603-2202, 1672-2241, 1698-2218, 1699-1973, 1733-2226, 1744-2226, 1745-2226, 1748-2222, 1748-2226, 1761-2218, 1764-2215, 1780-2215, 1784-2218, 1814-2218, 1827-2241, 1833-2226, 1853-2092, 1853-2174, 1853-2215, 1868-2241, 1897-2223, 2005-2215, 2007-2215, 2009-2215, 2138-2214
30/3176058CB1/ 1853	1-595, 4-520, 35-463, 37-181, 38-243, 41-325, 41-613, 41-727, 45-577, 75-509, 79-509, 148-863, 172-408, 173-417, 173-706, 210-649, 263-875, 319-857, 320-509, 408-761, 434-1081, 506-1109, 563-826, 597-783, 597-873, 597-878, 749-1161, 843-1089, 843-1305, 843-1359, 843-1621, 875-1223, 882-1400, 934-1390, 992-1244, 1071-1588, 1138-1687, 1175-1675, 1253-1838, 1263-1459, 1271-1400, 1273-1853, 1306-1448, 1307-1750, 1445-1478, 1448-1481, 1578-1610, 1578-1611, 1756-1801

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
31/2299818CB1/ 2541	1-271, 104-185, 139-393, 203-774, 292-905, 304-438, 343-438, 406-438, 409-649, 473-760, 473-788, 499-661, 503-661, 521-2541, 532-588, 551-588, 653-875, 653-890, 662-788, 662-1187, 667-890, 670-788, 891-1015, 891-1167, 903-1457, 919-1205, 924-1379, 963-1384, 970-1384, 973-1377, 986-1236, 986-1501, 1007-1526, 1021-1376, 1034-1255, 1034-1344, 1034-1479, 1039-1301, 1059-1659, 1078-1356, 1080-1347, 1131-1671, 1174-1434, 1203-1374, 1203-1437, 1270-2061, 1290-1516, 1307-1551, 1351-1633, 1351-1635, 1351-2046, 1406-1989, 1444-1862, 1501-2124, 1533-2109, 1533-2126, 1622-2170, 1631-1865, 1631-2092, 1637-2162, 1652-2200, 1683-1947, 1714-2293, 1718-2234, 1723-2172, 1736-2030, 1738-2320, 1750-1979, 1752-2445, 1767-2178, 1782-2001, 1823-2410, 1830-2471, 1832-2471, 1872-2413, 1872-2426, 1893-2382, 1893-2511, 1897-2309, 1906-2385, 1957-2485, 1975-2275, 1976-2478, 1979-2478, 1997-2238, 2048-2485, 2066-2288, 2080-2347, 2081-2524, 2093-2485, 2111-2470, 2111-2485, 2114-2522, 2115-2361, 2132-2485, 2154-2485, 2166-2409, 2180-2528, 2182-2485, 2275-2484, 2297-2528, 2336-2483, 2336-2486, 2410-2457
32/2729451CB1/ 4144	1-495, 1-578, 1-597, 1-623, 1-696, 2-419, 2-453, 2-606, 2-640, 10-471, 12-714, 23-408, 72-412, 124-398, 178-712, 188-473, 198-714, 254-958, 323-906, 357-862, 403-896, 471-697, 471-982, 655-1224, 739-853, 770-1030, 805-1349, 860-1089, 860-1331, 866-1451, 903-1483, 919-1150, 925-1487, 936-1213, 946-1424, 973-1751, 980-1495, 995-1729, 997-1519, 997-1635, 1030-1738, 1090-1728, 1095-1376, 1108-1205, 1143-1595, 1173-1855, 1204-3186, 1248-1870, 1328-1493, 1337-1906, 1343-1488, 1349-1976, 1349-1978, 1349-2056, 1351-1825, 1366-1943, 1374-2032, 1448-1753, 1460-1996, 1478-1935, 1486-2226, 1490-1721, 1508-1787, 1530-2325, 1593-2325, 1674-2325, 1698-1845, 1712-1914, 1729-2325, 1787-2325, 1810-2252, 1824-2325, 1846-2505, 1851-2517, 1863-2325, 1872-2325, 1885-2517, 1910-2325, 1991-2126, 2005-2573, 2005-2617, 2005-2629, 2073-2517, 2100-2264, 2100-2363, 2395-3004, 2434-3004, 2528-2753, 2528-3068, 2625-2835, 2800-3351, 2800-3462, 2876-3159, 3064-3534, 3149-3627, 3233-3505, 3235-3564, 3296-3551, 3296-3862, 3341-3802, 3431-4066, 3589-3851, 3872-4144

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
33/878534CB1/ 5218	1-566, 323-571, 323-720, 323-1039, 380-1054, 393-1007, 458-1059, 460-1155, 482-922, 485-1076, 520-766, 554-1041, 559-977, 641-1205, 850-1118, 853-1386, 1081-1354, 1135-1511, 1295-1900, 1401-1661, 1443-1935, 1488-1979, 1625-1818, 1634-2112, 1730-2393, 1814-2204, 1893-2503, 1929-2458, 1929-2543, 1991-2344, 1999-2693, 2050-2506, 2171-2516, 2230-2692, 2254-2565, 2255-2676, 2282-2726, 2355-2647, 2394-2567, 2401-2764, 2545-2831, 2658-2938, 2658-3193, 2784-2980, 2827-3414, 2899-3135, 2899-3184, 2958-3212, 3018-3823, 3070-3355, 3097-3364, 3192-3523, 3204-3448, 3273-3535, 3294-3543, 3319-3576, 3319-3817, 3333-3736, 3377-3668, 3391-3965, 3404-3736, 3517-3793, 3549-3814, 3604-3869, 3622-3861, 3637-3905, 3637-4172, 3638-4195, 3648-4025, 3648-4199, 3662-4200, 3694-3905, 3712-3933, 3713-3981, 3727-3919, 3729-3972, 3729-4246, 3827-4212, 3830-4102, 3875-4171, 3878-4123, 3883-4107, 3885-4171, 3902-4183, 3916-4186, 3968-4530, 3990-4264, 3995-4569, 4005-4285, 4051-4186, 4068-4363, 4069-4348, 4073-4351, 4075-4239, 4117-4368, 4144-4530, 4149-4425, 4187-4775, 4203-4364, 4203-4767, 4229-4758, 4231-4481, 4241-4391,
	4254-4528, 4256-4723, 4271-4524, 4271-4534, 4298-4560, 4301-4801, 4302-4785, 4331-4590, 4331-4824, 4374-4665, 4379-4982, 4465-4683, 4465-4996, 4495-4764, 4567-4707, 4595-4771, 4658-5218, 4690-4862, 4724-4958, 4724-5197, 4724-5214, 4725-5217, 4728-5027, 4772-5183, 4794-5217, 4798-5029, 4862-5093, 4913-5097
34/2806157CB1/ 763	1-602, 4-277, 16-613, 32-212, 32-486, 33-291, 34-310, 34-322, 34-330, 58-355, 58-649, 96-328, 100-346, 110-365, 473-763, 495-763
35/5883626CB1/ 869	1-150, 1-234, 1-250, 1-263, 1-272, 1-276, 1-278, 1-287, 8-202, 9-264, 15-590, 16-148, 17-320, 23-280, 23-594, 28-525, 34-288, 34-308, 63-325, 75-561, 128-721, 139-844, 182-439, 201-448, 203-662, 235-847, 279-326, 292-601, 292-744, 350-772, 411-855, 433-856, 460-856, 522-856, 526-852, 544-856, 551-856, 586-633, 590-633, 629-671, 629-677, 768-855, 768-869, 769-866

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
36/2674016CB1/ 2875	1-702, 9-245, 9-272, 9-525, 9-526, 9-547, 9-554, 9-626, 9-646, 11-459, 13-502, 13-597, 19-299, 19-393, 19-490, 23-102, 24-102, 24-219, 24-272, 24-273, 24-326, 25-102, 25-255, 25-270, 25-272, 25-273, 25-289, 25-292, 25-297, 25-305, 25-472, 25-480, 26-269, 26-292, 27-307, 29-321, 30-320, 31-102, 31-285, 31-448, 33-334, 34-102, 34-261, 34-308, 34-317, 34-381, 34-498, 34-538, 34-567, 36-282, 36-329, 36-338, 38-102, 39-319, 39-491, 41-248, 41-359, 45-322, 48-335, 48-354, 50-243, 50-529, 52-301, 52-336, 55-376, 57-276, 59-285, 59-307, 63-366, 64-337, 64-407, 73-102, 74-337, 90-375, 140-414, 144-720, 153-761, 182-345, 182-351, 182-353, 182-357, 182-370, 182-380, 182-733, 182-847, 182-857, 185-787, 186-389, 200-738, 214-879, 226-952, 236-661, 251-824, 254-783, 289-955, 290-817, 294-720, 309-935, 323-963, 339-1190, 347-932, 362-877, 387-663, 388-980, 402-935, 445-720, 455-890, 458-1044, 474-773, 476-963, 477-934, 493-1132, 494-1022, 496-763, 503-789, 504-959, 512-738, 516-764, 521-826, 526-550, 531-1048, 536-1011, 538-739, 538-1037, 539-845, 542-821, 542-993, 542-1005, 546-1153, 547-993, 550-1005, 557-1005, 564-1005, 575-993, 577-1211, 601-808,
	612-821, 632-767, 650-947, 650-1091, 654-939, 668-898, 671-933, 808-1061, 810-1202, 897-1132, 901-1142, 1012-1214, 1012-1232, 1012-1520, 1026-1215, 1026-1281, 1028-1702, 1099-1681, 1104-1696, 1148-1403, 1239-1710, 1246-1488, 1297-1509, 1299-1540, 1382-2042, 1384-1998, 1410-1702, 1410-1725, 1428-1981, 1455-2060, 1467-2043, 1476-1932, 1494-1989, 1499-1766, 1499-1767, 1499-1900, 1504-1969, 1542-2186, 1570-2214, 1611-2158, 1614-2208, 1632-2167, 1641-2085, 1653-1934, 1654-2224, 1661-1932, 1672-1960, 1673-1926, 1677-1697, 1689-2023, 1699-1959, 1705-2093, 1705-2147, 1711-1978, 1712-1992, 1713-2299, 1714-1978, 1714-1994, 1714-2010, 1714-2015, 1720-1960, 1722-1961, 1727-2050, 1735-2073, 1738-2016, 1738-2260, 1743-2000, 1743-2015, 1744-1990, 1745-2036, 1747-2254, 1748-1917, 1748-2253, 1749-2070, 1765-2024, 1772-1998, 1779-2105, 1799-1971, 1823-2095, 1828-2094, 1829-1919, 1831-2104, 1846-2118, 1846-2421, 1867-2112, 1881-2253, 1881-2290, 1884-2148, 1885-2150, 1889-2155, 1889-2201, 1893-2150, 1893-2159, 1893-2190, 1911-2165, 1915-2194, 1917-2302, 1918-2202, 1929-2296, 1931-2285, 1949-2242, 1973-2256, 1975-2283, 2028-2296, 2029-2300, 2032-2277,
	2045-2334, 2052-2289, 2069-2276, 2082-2233, 2086-2335, 2090-2673, 2110-2302, 2149-2220, 2163-2433, 2171-2434, 2171-2471, 2186-2875, 2207-2288, 2207-2417, 2214-2428, 2215-2442, 2216-2681, 2235-2553, 2242-2511, 2242-2710

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
37/5994159CB1/ 1839	1-291, 1-354, 1-449, 1-502, 67-485, 120-365, 189-725, 215-525, 215-570, 274-528, 331-915, 359-642, 376-914, 417-899, 526-1031, 539-699, 573-629, 605-1020, 628-782, 690-950, 692-1140, 727-1222, 731-1045, 778-1373, 791-1105, 820-1057, 829-1105, 841-1373, 856-1130, 856-1144, 856-1148, 919-1490, 933-1490, 934-1241, 1001-1315, 1012-1231, 1021-1490, 1021-1523, 1069-1587, 1140-1457, 1152-1573, 1172-1589, 1211-1816, 1235-1643, 1266-1769, 1269-1813, 1294-1709, 1336-1839, 1387-1592, 1391-1839, 1413-1839, 1538-1839, 1706-1826, 1707-1826
38/2457335CB1/ 1232	1-229, 1-271, 3-220, 3-224, 3-227, 3-240, 3-267, 22-278, 22-358, 25-260, 26-237, 26-259, 26-282, 26-387, 26-395, 26-489, 26-558, 26-576, 32-263, 32-289, 45-644, 47-292, 84-329, 116-374, 140-554, 350-616, 578-1167, 590-1014, 600-1151, 642-1232
39/2267802CB1/ 3250	1-479, 26-502, 34-539, 55-512, 58-496, 71-666, 79-842, 91-860, 116-705, 116-758, 116-760, 116-768, 116-775, 116-776, 116-779, 120-779, 123-779, 132-779, 142-779, 154-779, 159-779, 171-779, 212-779, 236-779, 265-779, 283-962, 300-537, 300-620, 300-779, 352-1000, 395-779, 404-779, 632-843, 632-1078, 644-1421, 651-1152, 921-1352, 939-1243, 976-1487, 1045-1238, 1048-1352, 1090-1305, 1262-1873, 1308-1796, 1371-2096, 1380-1890, 1465-1948, 1482-2113, 1483-2044, 1566-2019, 1628-2200, 1632-2175, 1771-2055, 1789-2352, 1799-2024, 1799-2101, 1799-2186, 1802-2040, 1867-2130, 1921-2123, 2083-2319, 2084-2369, 2084-2584, 2089-2326, 2156-2467, 2164-2565, 2168-2755, 2175-2455, 2175-2612, 2260-2497, 2260-2758, 2299-2583, 2301-2564, 2303-2485, 2443-2719, 2501-3153, 2508-2758, 2508-2998, 2510-2754, 2510-2778, 2538-2845, 2541-2706, 2550-3239, 2558-3207, 2611-2915, 2611-3206, 2611-3210, 2617-2892, 2630-3235, 2647-2948, 2666-2915, 2666-3234, 2670-3222, 2673-3234, 2690-3219, 2697-3231, 2730-2968, 2741-3194, 2745-3219, 2749-3016, 2754-3020, 2766-3236, 2796-3250, 2826-3052, 2836-3113, 2863-3162, 2864-3229,
	2882-3227, 2883-3241, 2885-3157, 2918-3213, 2918-3222, 2918-3250, 2935-3179, 2958-3237, 3046-3250, 3053-3241, 3064-3250, 3075-3238

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
40/3212060CB1/ 3621	1-317, 1-543, 1-591, 1-3602, 3-60, 14-296, 15-60, 17-60, 18-60, 18-662, 25-60, 28-60, 29-564, 238-995, 265-696, 310-602, 330-1001, 344-883, 382-822, 420-1260, 456-912, 458-1172, 461-888, 520-1093, 540-1151, 612-1287, 616-1100, 662-1158, 666-1390, 686-1373, 732-1433, 787-1380, 811-1400, 823-1447, 856-1430, 860-1386, 895-1223, 918-1287, 933-1409, 944-1408, 961-1576, 977-1636, 1015-1524, 1048-1602, 1052-1623, 1068-1614, 1098-1395, 1100-1758, 1159-3621, 1207-1813, 1215-1499, 1222-1517, 1243-1815, 1259-1664, 1316-1937, 1325-1852, 1331-1835, 1352-1967, 1356-2038, 1370-1784, 1407-1590, 1411-1984, 1436-1900, 1467-1725, 1493-2057, 1510-1978, 1588-2290, 1624-2161, 1645-2097, 1645-2268, 1648-1813, 1785-2391, 1811-2373, 1821-2311, 1834-2131, 1834-2143, 1843-2124, 1894-2346, 1897-2107, 1899-2159, 1899-2187, 1899-2288, 1899-2360, 1917-2468, 1934-2671, 1936-2276, 1961-2227, 1961-2391, 1996-2461, 2007-2315, 2026-2216, 2040-2691, 2046-2520, 2065-2391, 2098-2391, 2105-2375, 2114-2302, 2129-2623, 2129-2669, 2130-2449, 2153-2354, 2171-2391, 2211-2391, 2218-2391, 2235-2391, 2247-2893, 2256-2805, 2274-2523, 2276-2391, 2300-2548, 2306-2646, 2306-2845,
	2307-2748, 2313-2548, 2315-2391, 2348-2391, 2355-2391, 2359-2590, 2360-2826, 2391-2454, 2391-2521, 2391-2541, 2391-2590, 2391-2600, 2391-2656, 2391-2659, 2391-2760, 2391-2850, 2391-2885, 2391-2910, 2391-2932, 2391-2949, 2391-2960, 2391-2983, 2394-2970, 2395-3038, 2397-2600, 2411-2872, 2419-2952, 2420-3009, 2442-3034, 2449-3143, 2452-3021, 2458-2714, 2462-2639, 2503-3162, 2516-3054, 2526-2997, 2528-2978, 2531-2868, 2545-3054, 2550-2843, 2555-2865, 2674-2868, 2679-2971, 2684-3134, 2714-2977, 2732-2931, 2732-3288, 2741-3007, 2741-3152, 2805-3054, 2939-3137, 2945-3586, 2956-3544, 3011-3592, 3023-3295, 3025-3594, 3040-3485, 3089-3341, 3089-3580, 3089-3601, 3091-3361, 3127-3595, 3130-3602, 3136-3582, 3141-3601, 3151-3607, 3154-3414, 3155-3573, 3155-3608, 3158-3606, 3166-3609, 3167-3481, 3167-3503, 3167-3516, 3169-3600, 3170-3270, 3181-3604, 3185-3407, 3189-3608, 3207-3611, 3233-3621, 3249-3602, 3265-3601, 3292-3573, 3376-3602, 3399-3595, 3399-3621, 3401-3599, 3447-3601, 3490-3565, 3490-3621, 3513-3616
41/3121069CB1/ 1693	1-270, 1-475, 1-481, 1-507, 1-514, 3-475, 134-207, 134-319, 242-319, 385-994, 515-1030, 559-920, 690-1369, 711-1161, 792-1040, 792-1460, 794-1038, 860-1119, 894-1180, 1006-1693, 1115-1407
42/3280626CB1/ 2289	1-388, 2-120, 121-433, 286-507, 313-720, 434-1070, 596-881, 737-1363, 738-988, 738-1159, 738-1298, 741-1425, 773-1283, 802-975, 873-1558, 950-1700, 957-1485, 988-1293, 1046-1329, 1050-1384, 1096-1729, 1105-1767, 1135-1400, 1150-1958, 1168-1784, 1208-1668, 1219-1897, 1230-1823, 1295-1965, 1342-2065, 1368-2031, 1369-1741, 1387-1787, 1443-1670, 1501-1629, 1538-2146, 1567-2289

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
43/484404CB1/ 1304	1-641, 267-483, 269-311, 269-393, 269-400, 269-402, 269-455, 269-465, 269-486, 269-491, 269-500, 269-503, 269-509, 269-512, 269-519, 269-520, 269-523, 269-528, 269-531, 269-534, 269-537, 269-552, 269-555, 269-560, 269-561, 269-562, 269-568, 269-574, 269-585, 269-600, 269-613, 269-678, 269-688, 269-713, 269-746, 269-764, 269-769, 269-795, 269-810, 269-1287, 270-402, 270-478, 270-521, 270-534, 271-402, 271-746, 274-300, 275-516, 276-524, 276-525, 279-591, 287-522, 287-529, 287-550, 287-580, 287-603, 292-529, 294-402, 294-540, 296-678, 297-489, 311-494, 312-402, 312-496, 312-498, 312-505, 312-553, 312-575, 312-585, 312-677, 312-678, 312-795, 316-835, 328-603, 386-589, 404-465, 404-501, 404-607, 404-632, 404-639, 404-646, 404-960, 404-986, 448-939, 460-731, 466-717, 472-773, 489-678, 490-736, 492-727, 494-925, 508-780, 532-1122, 567-678, 571-678, 580-817, 583-916, 584-917, 585-857, 601-678, 608-1191, 640-862, 663-1216, 664-941, 674-946, 678-1087, 689-945, 700-1287, 716-1272, 716-1304, 717-941, 719-1002, 726-980, 729-940, 731-809, 731-812, 731-875, 731-884, 731-892, 731-907, 731-1203, 731-1276, 733-1260, 733-1285,
	736-790, 738-997, 739-987, 744-939, 744-984, 744-1030, 745-1004, 745-1014, 754-946, 754-964, 771-947, 771-1039, 783-1038, 783-1066, 783-1087, 791-1082, 791-1213, 797-1302, 800-1094, 807-1287, 810-1289, 814-980, 824-1082, 824-1083, 831-1280, 831-1287, 836-1137, 839-1287, 839-1294, 840-1002, 847-1301, 848-1286, 850-1038, 850-1105, 851-1190, 855-1205, 856-1109, 860-1287, 879-1151, 897-919, 917-1165, 920-1188, 924-1225, 935-1037, 935-1292, 973-1291, 1035-1304, 1036-1236, 1036-1275, 1049-1302, 1071-1287, 1113-1287, 1165-1287, 1171-1287
44/2830063CB1/ 4850	1-584, 92-315, 95-350, 95-473, 95-475, 101-742, 104-332, 106-590, 108-475, 130-487, 158-338, 234-501, 234-502, 248-468, 312-473, 312-492, 312-711, 312-802, 351-803, 386-407, 444-1145, 462-1105, 624-910, 712-1029, 789-1066, 789-1387, 873-1153, 913-1142, 963-1174, 1019-1696, 1069-1712, 1074-1508, 1103-1500, 1178-1393, 1178-1714, 1178-1808, 1200-2046, 1225-1296, 1225-1318, 1225-1436, 1225-1439, 1225-1600, 1277-1592, 1379-1598, 1440-1504, 1440-1610, 1533-1792, 1533-2010, 1574-2151, 1581-1836, 1581-2009, 1600-1854, 1600-1891, 1632-1909, 1687-2117, 1695-2170, 1696-1938, 1761-2351, 1817-2345, 1817-2363, 1818-2344, 1865-2459, 1905-2446, 1914-2588, 1918-2038, 1972-2282, 2017-2637, 2021-2468, 2066-2271, 2100-2478, 2137-2536, 2153-2493, 2205-2802, 2217-2613, 2237-2478, 2237-2494, 2237-2514, 2238-2605, 2247-2759, 2276-2635, 2280-2592, 2293-2427, 2306-2534, 2318-2893, 2327-2843, 2329-2581, 2329-2624, 2329-2648, 2338-2776, 2370-3043, 2482-2787, 2536-2800, 2567-2808, 2617-3235,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
44	2617-3405, 2627-2845, 2627-2898, 2627-2959, 2627-3078, 2627-3283, 2627-3306, 2627-3310, 2627-3315, 2627-3339, 2627-3348, 2627-3369, 2627-3372, 2627-3405, 2629-3374, 2630-3009, 2632-2769, 2773-3593, 2803-3450, 2809-2978, 2809-3341, 2811-3305, 2830-3688, 2834-3271, 2837-3678, 2847-3607, 2903-3067, 2903-3305, 2909-3212, 2909-3503, 2910-3626, 2917-3545, 2924-3727, 2930-3271, 2958-3543, 2971-3582, 2971-3588, 2987-3740, 2987-3756, 2993-3767, 2996-3271, 3018-3674, 3023-3861, 3024-3723, 3031-3873, 3041-3271, 3047-3759, 3048-3319, 3053-3390, 3064-3695, 3066-3332, 3067-3855, 3067-3908, 3078-3477, 3082-3789, 3084-3683, 3095-3919, 3114-3741, 3139-3838, 3147-3881, 3152-3713, 3156-3684, 3163-3688, 3170-3750, 3183-3776, 3190-3666, 3191-3792, 3195-3666, 3195-3769, 3207-3827, 3210-3864, 3211-4068, 3212-3545, 3227-3778, 3237-3469, 3241-3833, 3244-3984, 3248-3991, 3259-3674, 3265-3875, 3268-4010, 3278-4120, 3282-3914, 3296-4016, 3296-4122, 3298-4164, 3300-4135, 3301-3849, 3315-3582, 3315-3606, 3318-3991, 3319-4158, 3320-3982, 3326-3982, 3336-3911, 3337-3953, 3343-3585, 3345-3958, 3347-3818, 3350-3944, 3352-4282, 3375-4014, 3375-4046, 3380-4174,
	3401-4034, 3401-4039, 3435-4106, 3435-4149, 3439-4134, 3454-3984, 3465-4166, 3474-4062, 3476-4238, 3481-4096, 3481-4169, 3482-4062, 3482-4320, 3495-4191, 3531-4155, 3531-4197, 3531-4251, 3531-4269, 3531-4307, 3531-4319, 3531-4382, 3532-4312, 3536-4277, 3563-4200, 3566-4190, 3566-4204, 3569-3841, 3573-3613, 3590-4295, 3592-3926, 3593-4098, 3606-4429, 3643-4447, 3656-4422, 3663-4346, 3668-4372, 3678-4318, 3688-4376, 3694-4470, 3713-4510, 3722-4377, 3729-4587, 3730-4569, 3739-4370, 3740-4329, 3746-4379, 3752-4355, 3759-4435, 3762-4595, 3762-4602, 3773-4518, 3787-4427, 3787-4479, 3796-3857, 3797-4652, 3805-4571, 3821-4651, 3831-4503, 3833-4690, 3841-4376, 3845-4678, 3854-4489, 3856-4414, 3856-4729, 3858-4585, 3865-4327, 3865-4329, 3865-4426, 3865-4433, 3865-4471, 3865-4524, 3865-4531, 3865-4538, 3865-4578, 3865-4582, 3865-4652, 3865-4674, 3865-4680, 3865-4688, 3868-4548, 3868-4550, 3868-4758, 3872-4459, 3876-4496, 3877-4578, 3886-4424, 3889-4738, 3891-4456, 3891-4628, 3893-4306, 3912-4777, 3926-4448, 3926-4713, 3926-4834, 3929-4834, 3933-4448, 3938-4678, 3949-4547, 3958-4834, 3960-4448, 3962-4452, 3982-4414, 3989-4834, 4000-4622,
	4012-4414, 4013-4521, 4018-4834, 4024-4471, 4032-4834, 4035-4834, 4037-4834, 4044-4693, 4044-4843, 4048-4310, 4048-4414, 4099-4827, 4144-4846, 4146-4850, 4167-4845, 4178-4657, 4187-4464, 4187-4673, 4201-4820, 4246-4813, 4252-4549, 4273-4537, 4284-4733

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
45/7506096CB1/ 4350	1-4346, 1-4350, 425-841, 502-943, 503-937, 525-841, 530-943, 539-943, 650-943, 666-923, 1580-1905, 2124-2385, 2125-2674, 2125-2707, 2409-2628, 2409-2881, 2409-3115, 2614-3257, 2635-2941, 2635-3036, 2635-3091, 2667-3154, 2677-3281, 2681-3262, 2734-3273, 2739-3281, 2926-3081, 3015-3552, 3041-3396, 3742-4316, 3819-4350, 3824-4350, 3886-4350, 3931-4348, 3931-4350, 3932-4349, 3932-4350, 3983-4345, 4109-4284, 4131-4350, 4181-4284
46/7505914CB1/ 2959	1-702, 2-2822, 9-245, 9-272, 9-526, 9-547, 11-459, 13-597, 16-832, 19-299, 24-219, 24-271, 24-272, 24-273, 24-326, 24-619, 25-199, 25-204, 25-222, 25-270, 25-272, 25-273, 25-292, 25-305, 25-317, 25-431, 25-434, 25-647, 25-735, 25-787, 26-220, 26-255, 26-269, 26-289, 26-292, 26-297, 27-304, 27-307, 29-321, 30-298, 30-320, 30-693, 31-285, 31-448, 33-334, 34-255, 34-261, 34-288, 34-291, 34-308, 34-317, 34-349, 34-492, 34-505, 34-538, 34-567, 34-575, 34-635, 34-677, 34-911, 35-206, 35-272, 35-364, 36-282, 36-329, 36-338, 36-622, 36-658, 37-295, 37-301, 37-322, 37-343, 39-276, 39-296, 39-319, 39-491, 39-600, 39-656, 39-659, 39-701, 40-296, 41-169, 41-248, 41-274, 41-300, 41-359, 41-531, 41-557, 41-559,
	41-813, 43-537, 46-295, 46-308, 47-322, 48-354, 49-335, 50-243, 50-324, 50-405, 50-529, 51-406, 52-287, 52-308, 52-311, 52-325, 53-301, 53-336, 53-344, 54-361, 55-376, 57-276, 58-865, 59-285, 59-307, 59-337, 59-886, 62-605, 63-366, 64-407, 65-232, 65-336, 65-383, 65-654, 70-581, 75-337, 76-318, 83-368, 87-634, 89-313, 90-375, 95-865, 116-715, 140-414, 140-449, 144-720, 153-761, 176-463, 184-462, 185-787, 200-735, 214-871, 218-484, 218-866, 219-459, 221-501, 223-490, 226-952, 234-475, 236-661, 243-517, 245-1009, 246-964, 251-517, 254-783, 259-596, 265-963, 268-496, 289-955, 290-817, 293-550, 294-720, 319-621, 323-963, 339-1190, 343-653, 347-932, 362-877, 363-594, 372-623, 374-619, 380-883, 387-663, 402-935, 414-702, 445-720, 455-890, 458-1044, 474-773, 475-1016, 476-959, 477-934,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
46	494-1022, 496-763, 503-789, 512-738, 512-994, 516-764, 521-826, 536-1003, 537-1004, 538-739, 538-1037, 540-845, 542-821, 542-993, 542-1005, 543-992, 546-1153, 547-993, 550-1005, 550-1171, 557-1005, 564-1005, 569-1092, 575-993, 577-1211, 579-993, 598-1005, 599-1018, 602-808, 602-1155, 607-880, 612-821, 612-1194, 613-872, 617-1011, 621-1003, 624-1247, 632-767, 640-1242, 641-1005, 644-1167, 650-947, 650-1091, 653-1019, 654-939, 666-1073, 669-898, 671-933, 676-980, 684-1005, 685-1262, 687-1240, 702-994, 714-1005, 721-1005, 739-873, 742-1019, 743-1019, 748-867, 748-959, 764-1069, 766-993, 768-846, 772-1113, 802-1005, 810-1061, 810-1202, 873-1152, 934-1011, 936-1211, 947-1198, 949-1233, 957-1272, 958-1221, 960-1121, 976-1271, 981-1264, 993-1181, 1006-1283, 1012-1214, 1012-1232, 1014-1285, 1026-1215, 1026-1281, 1034-1299, 1049-1310, 1058-1303, 1083-1287, 1095-1268,
	1272-1656, 1277-1659, 1316-1539, 1328-1988, 1330-1944, 1340-1739, 1356-1634, 1361-1648, 1366-1633, 1397-1705, 1401-2006, 1418-1682, 1422-1878, 1437-1615, 1440-1935, 1443-1711, 1445-1712, 1445-1713, 1445-1846, 1450-1915, 1471-1790, 1480-1659, 1487-1757, 1488-2132, 1495-2239, 1502-1746, 1502-1818, 1516-2160, 1527-1831, 1539-1810, 1544-1774, 1548-1714, 1553-1829, 1553-1835, 1560-2154, 1570-1808, 1577-2226, 1580-2113, 1582-1825, 1582-2234, 1584-1802, 1587-2031, 1599-1880, 1600-2170, 1607-1790, 1613-1829, 1618-1906, 1619-1872, 1635-1969, 1645-1905, 1651-2039, 1651-2093, 1657-1924, 1659-2147, 1660-1840, 1660-1924, 1660-1940, 1660-1956, 1660-1961, 1662-1938, 1666-1806, 1666-1906, 1668-1907, 1675-1996, 1681-2019, 1684-1962, 1686-2199, 1689-1946, 1690-1936, 1690-1961,
	1691-1982, 1693-2200, 1694-1863, 1694-2199, 1695-2016, 1715-1970, 1718-1944, 1725-2051, 1745-1917, 1769-2041, 1774-2040, 1777-2050, 1792-2064, 1792-2367, 1813-2058, 1827-2199, 1827-2236, 1830-2094, 1831-1878, 1831-2096, 1835-2101, 1835-2147, 1839-2136, 1840-2096, 1842-2105, 1857-2111, 1861-2140, 1866-2148, 1875-2242, 1877-2231, 1882-2130, 1895-2188, 1897-2139, 1903-2142, 1919-2202, 1923-2229, 1974-2242, 1975-2246, 1978-2223, 1986-2258, 1991-2280, 1998-2235, 2026-2146, 2028-2138, 2028-2179, 2036-2619, 2056-2248, 2085-2204, 2109-2379, 2117-2417, 2132-2822, 2153-2363, 2157-2812, 2160-2374, 2162-2627, 2188-2656, 2189-2586, 2190-2457, 2198-2416, 2198-2437, 2198-2496, 2199-2463, 2206-2481, 2210-2690, 2211-2461, 2211-2573, 2213-2479, 2215-2463, 2216-2452, 2218-2516, 2220-2515, 2223-2793, 2223-2801, 2224-2428,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
46	2225-2478, 2225-2542, 2228-2500, 2233-2513, 2233-2521, 2234-2496, 2235-2481, 2238-2494, 2240-2429, 2240-2690, 2243-2492, 2243-2558, 2251-2548, 2256-2506, 2258-2725, 2261-2540, 2263-2495, 2265-2514, 2265-2538, 2290-2604, 2291-2533, 2292-2512, 2302-2876, 2304-2831, 2333-2567, 2333-2606, 2344-2482, 2344-2608, 2356-2637, 2357-2602, 2363-2627, 2371-2627, 2380-2634, 2382-2573, 2382-2601, 2382-2616, 2382-2629, 2384-2677, 2401-2653, 2413-2689, 2418-2669, 2431-2701, 2434-2897, 2437-2666, 2437-2679, 2437-2709, 2438-2704, 2438-2709, 2445-2710, 2448-2617, 2453-2703, 2454-2723, 2468-2719, 2478-2776, 2480-2777, 2513-2824, 2515-2791, 2533-2959, 2535-2742, 2537-2827, 2539-2801, 2550-2794, 2551-2797, 2553-2841, 2562-2842, 2573-2798, 2582-2824, 2593-2780, 2623-2759, 2637-2843, 2672-2776

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID:	Representative Library
24	71230017CB1	LUNGNOT35
25	3125036CB1	LIVRNON08
26	1758089CB1	BRAITDR03
27	3533891CB1	HELATXT05
28	1510943CB1	OVARTUE01
29	2119377CB1	PANCNOT05
30	3176058CB1	ADRENON04
31	2299818CB1	BRABDIR01
32	2729451CB1	PROSNON01
33	878534CB1	PITUNOT03
34	2806157CB1	BLADTUT08
35	5883626CB1	LIVRNON08
36	2674016CB1	BEPINOT01
37	5994159CB1	SKINNOT05
38	2457335CB1	ENDANOT01
39	2267802CB1	EPIPNOT01
40	3212060CB1	THYMNOT08
41	3121069CB1	COLNTUT02
42	3280626CB1	STOMFET02
43	484404CB1	PROSTUT09
44	2830063CB1	TLYMNOT03
45	7506096CB1	TLYMNOT05
46	7505914CB1	TLYMTXT02

Table 6

Library	Vector	Library Description
ADRENON04	PSPORT1	Normalized library was constructed from 1.36 x 1e6 independent clones from an adrenal tissue library. Starting RNA was made from adrenal gland tissue removed from a 20-year-old Caucasian male, who died from head trauma. The library was normalized in two rounds using conditions adapted from Soares et al. (PNAS (1994) 91:9228-9232) and Bonaldo et al. (Genome Res (1996) 6: 791-806), using a significantly longer (48-hours/round) reannealing hybridization period.
BEPINOT01	PSPORT1	Library was constructed using RNA isolated from a bronchial epithelium primary cell line derived from a 54-year-old Caucasian male.
BLADTUT08	pINCY	Library was constructed using RNA isolated from bladder tumor tissue removed from a 72-year-old Caucasian male during a radical cystectomy and prostatectomy. Pathology indicated an invasive grade 3 (of 3) transitional cell carcinoma in the right bladder base. Patient history included pure hypercholesterolemia and tobacco abuse. Family history included myocardial infarction, cerebrovascular disease, and brain cancer.
BRABDIR01	pINCY	Library was constructed using RNA isolated from diseased cerebellum tissue removed from the brain of a 57-year-old Caucasian male, who died from a cerebrovascular accident. Patient history included Huntington's disease, emphysema, and tobacco abuse.
BRAITDR03	PCDNA2.1	This random primed library was constructed using RNA isolated from allocortex, cingulate posterior tissue removed from a 55-year-old Caucasian female who died from cholangiocarcinoma. Pathology indicated mild meningeal fibrosis predominately over the convexities, scattered axonal spheroids in the white matter of the cingulate cortex and the thalamus, and a few scattered neurofibrillary tangles in the entorhinal cortex and the periaqueductal gray region. Pathology for the associated tumor tissue indicated well-differentiated cholangiocarcinoma of the liver with residual or relapsed tumor. Patient history included cholangiocarcinoma, post-operative Budd-Chiari syndrome, biliary ascites, hydrothorax, dehydration, malnutrition, oliguria and acute renal failure. Previous surgeries included cholecystectomy and resection of 85% of the liver.
COLNTUT02	PSPORT1	Library was constructed using RNA isolated from colon tumor tissue removed from a 75-year-old Caucasian male during a hemicolectomy. Pathology indicated invasive grade 3 adenocarcinoma arising in a tubulovillous adenoma, which was distal to the ileocecal valve in the cecum. The tumor penetrated deeply into the muscularis propria but not through it.
ENDANOT01	PBLUESCRIPT	Library was constructed using RNA isolated from aortic endothelial cell tissue from an explanted heart removed from a male during a heart transplant.
EPIPNOT01	pINCY	Library was constructed using RNA isolated from prostatic epithelial cells removed from a 17-year-old Hispanic male.

Table 6

Library	Vector	Library Description
HELATXT05	pINCY	Library was constructed using RNA isolated from a treated HeLa cell line, derived from cervical adenocarcinoma removed from a 31-year-old Black female. The cells were treated with 25 microM sodium butyrate for 24 hours.
LIVRNON08	pINCY	This normalized library was constructed from 5.7 million independent clones from a pooled liver tissue library. Starting RNA was made from pooled liver tissue removed from a 4-year-old Hispanic male who died from anoxia and a 16 week female fetus who died after 16-weeks gestation from anencephaly. Serologies were positive for cytolomegalovirus in the 4-year-old. Patient history included asthma in the 4-year-old. Family history included taking daily prenatal vitamins and mitral valve prolapse in the mother of the fetus. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228 and Bonaldo et al., Genome Research 6 (1996):791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
LUNGNOT35	pINCY	Library was constructed using RNA isolated from lung tissue removed from a 62-year-old Caucasian female. Pathology for the associated tumor tissue indicated a grade 1 spindle cell carcinoid forming a nodule. Patient history included depression, thrombophlebitis, and hyperlipidemia. Family history included cerebrovascular disease, atherosclerotic coronary artery disease, breast cancer, colon cancer,type II diabetes, and malignant skin melanoma.
OVARTUE01	PCDNA2.1	This 5' biased random primed library was constructed using RNA isolated from left ovary tumor tissue removed from a 44-year-old female. Pathology indicated grade 4 (of 4) serous carcinoma replacing both the right and left ovaries forming solid mass cystic masses. Neoplastic deposits were identified in para-ovarian soft tissue.
PANCNOT05	PSPORT1	Library was constructed using RNA isolated from the pancreatic tissue of a 2-year-old Hispanic male who died from cerebral anoxia.
PITUNOT03	PSPORT1	Library was constructed using RNA isolated from pituitary tissue of a 46-year-old Caucasian male, who died from colon cancer. Serologies were negative. Patient history included arthritis, peptic ulcer disease, and tobacco use. Patient medications included Tagamet and muscle relaxants.
PROSNON01	PSPORT1	This normalized prostate library was constructed from 4.4 M independent clones from a prostate library. Starting RNA was made from prostate tissue removed from a 28-year-old Caucasian male who died from a self-inflicted gunshot wound. The normalization and hybridization conditions were adapted from Soares, M.B. et al. (1994) Proc. Natl. Acad. Sci. USA 91:9228-9232, using a longer (19 hour) reannealing hybridization period.

Table 6

Library	Vector	Library Description
PROSTUT09	pINCY	Library was constructed using RNA isolated from prostate tumor tissue removed from a 66-year-old Caucasian male during a radical prostatectomy, radical cystectomy, and urinary diversion. Pathology indicated grade 3 transitional cell carcinoma. The patient presented with prostatic inflammatory disease. Patient history included lung neoplasm, and benign hypertension. Family history included a malignant breast neoplasm, tuberculosis, cerebrovascular disease, atherosclerotic coronary artery disease and lung cancer.
SKINNOT05	pINCY	Library was constructed using RNA isolated from skin tissue removed from a Caucasian male fetus, who died from Patau's syndrome (trisomy 13) at 20-weeks' gestation.
STOMFET02	pINCY	Library was constructed using RNA isolated from stomach tissue removed from a Hispanic male fetus, who died at 18 weeks' gestation.
THYMNOT08	pINCY	Library was constructed using RNA isolated from thymus tissue removed from a 4-month-old Caucasian male during a total thymectomy and open heart repair of atrioventricular canal defect using hypothermia. Pathology indicated a grossly normal thymus. The patient presented with a congenital heart anomaly, congestive heart failure, and Down's syndrome. Patient history included abnormal thyroid function study and premature birth. Previous procedures included right and left heart angiocardiography. Patient medications included Digoxin, Synthroid, and Lasix.
TLYMNOT03	pINCY	Library was constructed using RNA isolated from nonactivated Th1 cells. These cells were differentiated from umbilical cord CD4 T cells with IL-12 and B7-transfected COS cells.
TLYMNOT05	pINCY	Library was constructed using RNA isolated from nonactivated Th2 cells. These cells were differentiated from umbilical cord CD4 T cells with IL-4 in the presence of anti-IL-12 antibodies and B7-transfected COS cells.
TLYMTXT02	pINCY	Library was constructed using RNA isolated from CD4+ T cells obtained from a pool of donors. The cells were treated with CD3 antibodies.

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	*ESTs:* Probability value=1.0E-8 or less *Full Length sequences:* Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	*ESTs:* fasta E value=1.06E-6 *Assembled ESTs:* fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less *Full Length sequences:* fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value=1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM, INCY, SMART, and TIGRFAM.	Krogh, A. et al. (1994) J. Mol. Biol., 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	*PFAM, INCY, SMART, or TIGRFAM hits:* Probability value=1.0E-3 or less *Signal peptide hits:* Score= 0 or greater

Table 7 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score≥GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score=120 or greater; Match length=56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

PF-#### P

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
23	7505914	1243473H1	SNP00053642	20	916	T	T	G	W235	n/a	n/a	n/a	n/a
23	7505914	1378379H1	SNP00053642	107	916	T	T	G	W235	n/a	n/a	n/a	n/a
23	7505914	1892720H1	SNP00053642	246	916	T	T	G	W235	n/a	n/a	n/a	n/a
23	7505914	2278476H1	SNP00053642	263	916	T	T	G	W235	n/a	n/a	n/a	n/a
23	7505914	2926120H1	SNP00053642	148	913	T	T	G	S234	n/a	n/a	n/a	n/a
23	7505914	3032515H1	SNP00053642	141	912	T	T	G	P233	n/a	n/a	n/a	n/a
23	7505914	3249576H1	SNP00144526	24	2740	A	A	G	noncoding	n/a	n/a	n/a	n/a
23	7505914	3902996H1	SNP00144526	19	2740	A	A	G	noncoding	n/a	n/a	n/a	n/a
23	7505914	4723089H1	SNP00144526	23	2739	A	A	G	noncoding	n/a	n/a	n/a	n/a
23	7505914	6199516H1	SNP00053642	423	916	T	T	G	W235	n/a	n/a	n/a	n/a
23	7505914	6266935H1	SNP00140788	169	2719	T	T	C	noncoding	n/a	n/a	n/a	n/a
23	7505914	6266935H1	SNP00144526	190	2740	A	A	G	noncoding	n/a	n/a	n/a	n/a
23	7505914	6588185H1	SNP00053642	316	916	T	T	G	W235	n/a	n/a	n/a	n/a
23	7505914	6830234H1	SNP00053642	305	916	G	T	G	G235	n/a	n/a	n/a	n/a
23	7505914	6990628H1	SNP00144526	7	2740	A	A	G	noncoding	n/a	n/a	n/a	n/a
23	7505914	7428572H1	SNP00053642	370	916	T	T	G	W235	n/a	n/a	n/a	n/a
23	7505914	7684829H1	SNP00140788	219	2719	C	T	C	noncoding	n/a	n/a	n/a	n/a
23	7505914	7684829H1	SNP00144526	240	2740	G	A	G	noncoding	n/a	n/a	n/a	n/a
23	7505914	8618556J1	SNP00053642	611	916	T	T	G	W235	n/a	n/a	n/a	n/a

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-23,

b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, SEQ ID NO:17, and SEQ ID NO:19-23,

c) a naturally occurring polypeptide comprising an amino acid sequence at least 99% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:15-16 and SEQ ID NO:18,

d) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, and

e) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23.

2. An isolated polypeptide of claim 1 comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-23.

3. An isolated polynucleotide encoding a polypeptide of claim 1.

4. An isolated polynucleotide encoding a polypeptide of claim 2.

5. An isolated polynucleotide of claim 4 comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:24-46.

6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.

7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.

9. A method of producing a polypeptide of claim 1, the method comprising:

a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and

b) recovering the polypeptide so expressed.

10. A method of claim 9, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-23.

11. An isolated antibody which specifically binds to a polypeptide of claim 1.

12. An isolated polynucleotide selected from the group consisting of:

a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:24-46,

b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:24-46,

c) a polynucleotide complementary to a polynucleotide of a),

d) a polynucleotide complementary to a polynucleotide of b), and

e) an RNA equivalent of a)-d).

13. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 12.

14. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and

b) detecting the presence or absence of said hybridization complex, and, optionally, if

present, the amount thereof.

15. A method of claim 14, wherein the probe comprises at least 60 contiguous nucleotides.

16. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and

b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

17. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

18. A composition of claim 17, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-23.

19. A method for treating a disease or condition associated with decreased expression of functional MDDT, comprising administering to a patient in need of such treatment the composition of claim 17.

20. A method of screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting agonist activity in the sample.

21. A composition comprising an agonist compound identified by a method of claim 20 and a pharmaceutically acceptable excipient.

22. A method for treating a disease or condition associated with decreased expression of functional MDDT, comprising administering to a patient in need of such treatment a composition of claim 21.

23. A method of screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting antagonist activity in the sample.

24. A composition comprising an antagonist compound identified by a method of claim 23 and a pharmaceutically acceptable excipient.

25. A method for treating a disease or condition associated with overexpression of functional MDDT, comprising administering to a patient in need of such treatment a composition of claim 24.

26. A method of screening for a compound that specifically binds to the polypeptide of claim 1, the method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and

b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

27. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, the method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,

b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and

c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

28. A method of screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,

b) detecting altered expression of the target polynucleotide, and

c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

29. A method of assessing toxicity of a test compound, the method comprising:

a) treating a biological sample containing nucleic acids with the test compound,

b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 12 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 12 or fragment thereof,

c) quantifying the amount of hybridization complex, and

d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

30. A diagnostic test for a condition or disease associated with the expression of MDDT in a biological sample, the method comprising:

a) combining the biological sample with an antibody of claim 11, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex, and

b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

31. The antibody of claim 11, wherein the antibody is:

a) a chimeric antibody,

b) a single chain antibody,

c) a Fab fragment,

d) a $F(ab')_2$ fragment, or

e) a humanized antibody.

32. A composition comprising an antibody of claim 11 and an acceptable excipient.

33. A method of diagnosing a condition or disease associated with the expression of MDDT in a subject, comprising administering to said subject an effective amount of the composition of claim 32.

34. A composition of claim 32, wherein the antibody is labeled.

35. A method of diagnosing a condition or disease associated with the expression of MDDT in a subject, comprising administering to said subject an effective amount of the composition of claim 34.

36. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 11, the method comprising:

a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, or an immunogenic fragment thereof, under conditions to elicit an antibody response,

b) isolating antibodies from said animal, and

c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-23.

37. A polyclonal antibody produced by a method of claim 36.

38. A composition comprising the polyclonal antibody of claim 37 and a suitable carrier.

39. A method of making a monoclonal antibody with the specificity of the antibody of claim 11, the method comprising:

a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, or an immunogenic fragment thereof, under conditions to elicit an antibody response,

b) isolating antibody producing cells from the animal,

c) fusing the antibody producing cells with immortalized cells to form monoclonal

antibody-producing hybridoma cells,

d) culturing the hybridoma cells, and

e) isolating from the culture monoclonal antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-23.

40. A monoclonal antibody produced by a method of claim 39.

41. A composition comprising the monoclonal antibody of claim 40 and a suitable carrier.

42. The antibody of claim 11, wherein the antibody is produced by screening a Fab expression library.

43. The antibody of claim 11, wherein the antibody is produced by screening a recombinant immunoglobulin library.

44. A method of detecting a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-23 in a sample, the method comprising:

a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and

b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-23 in the sample.

45. A method of purifying a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-23 from a sample, the method comprising:

a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and

b) separating the antibody from the sample and obtaining the purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-23.

46. A microarray wherein at least one element of the microarray is a polynucleotide of claim

13.

47. A method of generating an expression profile of a sample which contains polynucleotides, the method comprising:

a) labeling the polynucleotides of the sample,

b) contacting the elements of the microarray of claim 46 with the labeled polynucleotides of the sample under conditions suitable for the formation of a hybridization complex, and

c) quantifying the expression of the polynucleotides in the sample.

48. An array comprising different nucleotide molecules affixed in distinct physical locations on a solid substrate, wherein at least one of said nucleotide molecules comprises a first oligonucleotide or polynucleotide sequence specifically hybridizable with at least 30 contiguous nucleotides of a target polynucleotide, and wherein said target polynucleotide is a polynucleotide of claim 12.

49. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 30 contiguous nucleotides of said target polynucleotide.

50. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 60 contiguous nucleotides of said target polynucleotide.

51. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to said target polynucleotide.

52. An array of claim 48, which is a microarray.

53. An array of claim 48, further comprising said target polynucleotide hybridized to a nucleotide molecule comprising said first oligonucleotide or polynucleotide sequence.

54. An array of claim 48, wherein a linker joins at least one of said nucleotide molecules to said solid substrate.

55. An array of claim 48, wherein each distinct physical location on the substrate contains

multiple nucleotide molecules, and the multiple nucleotide molecules at any single distinct physical location have the same sequence, and each distinct physical location on the substrate contains nucleotide molecules having a sequence which differs from the sequence of nucleotide molecules at another distinct physical location on the substrate.

56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

60. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

61. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

62. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.

63. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.

64. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.

65. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:10.

66. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:11.

67. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:12.

68. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:13.

69. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:14.

70. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:15.

71. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:16.

72. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:17.

73. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:18.

74. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:19.

75. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:20.

76. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:21.

77. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:22.

78. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:23.

79. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:24.

80. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:25.

81. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:26.

82. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:27.

83. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:28.

84. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:29.

85. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:30.

86. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:31.

87. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:32.

88. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:33.

89. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:34.

90. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:35.

91. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:36.

92. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:37.

93. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:38.

94. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:39.

95. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:40.

96. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:41.

97. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:42.

98. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:43.

99. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:44.

100. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:45.

101. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:46.

<110> INCYTE GENOMICS, INC.
LU, DYUNG AINA M.
ARVIZU, CHANDRA S.
GANDHI, AMEENA R.
HAFALIA, APRIL J.A.
DING, LI
LU, YAN
RAMKUMAR, JAYALAXMI
SWARNAKER, ANITA
TANG, Y. TOM
YUE, HENRY
TRAN, BAO
LEE, SOO Y.
WARREN, BRIDGET A.
NGUYEN, DANNIEL B.
THANGAVELU, KAVITHA
YAO, MONIQUE G.
ELLIOTT, VICKI S.
BAUGHN, MARIAH R.
EMERLING, BROOKE M.
LAL, PREETI
GIETZEN, KIMBERLY J.
BECHA, SHANYA D.
MARQUIS, JOSEPH P.
KABLE, AMY E.

<120> MOLECULES FOR DISEASE DETECTION AND TREATMENT

<130> PF-0921 PCT

<140> To Be Assigned
<141> Herewith

<150> 60/280,387; 60/282,335; 60/286,663; 60/285,484; 60/350,702; 60/351,749
<151> 2001-03-30; 2001-04-05; 2001-04-13; 2001-04-19; 2002-01-18; 2002-01-25

<160> 46
<170> PERL Program

<210> 1
<211> 485
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 71230017CD1

<400> 1
Met Asp Pro Thr Ala Leu Val Glu Ala Ile Val Glu Glu Val Ala
1 5 10 15
Cys Pro Ile Cys Met Thr Phe Leu Arg Glu Pro Met Ser Ile Asp
20 25 30
Cys Gly His Ser Phe Cys His Ser Cys Leu Ser Gly Leu Trp Glu
35 40 45

Ile Pro Gly Glu Ser Gln Asn Trp Gly Tyr Thr Cys Pro Leu Cys
50 55 60
Arg Ala Pro Val Gln Pro Arg Asn Leu Arg Pro Asn Trp Gln Leu
65 70 75
Ala Asn Val Val Glu Lys Val Arg Leu Leu Arg Leu His Pro Gly
80 85 90
Met Gly Leu Lys Gly Asp Leu Cys Glu Arg His Gly Glu Lys Leu
95 100 105
Lys Met Phe Cys Lys Glu Asp Val Leu Ile Met Cys Glu Ala Cys
110 115 120
Ser Gln Ser Pro Glu His Glu Ala His Ser Val Val Pro Met Glu
125 130 135
Asp Val Ala Trp Glu Tyr Lys Trp Glu Leu His Glu Ala Leu Glu
140 145 150
His Leu Lys Lys Glu Gln Glu Glu Ala Trp Lys Leu Glu Val Gly
155 160 165
Glu Arg Lys Arg Thr Ala Thr Trp Lys Ile Gln Val Glu Thr Arg
170 175 180
Lys Gln Ser Ile Val Trp Glu Phe Glu Lys Tyr Gln Arg Leu Leu
185 190 195
Glu Lys Lys Gln Pro Pro His Arg Gln Leu Gly Ala Glu Val Ala
200 205 210
Ala Ala Leu Ala Ser Leu Gln Arg Glu Ala Ala Glu Thr Met Gln
215 220 225
Lys Leu Glu Leu Asn His Ser Glu Leu Ile Gln Gln Ser Gln Val
230 235 240
Leu Trp Arg Met Ile Ala Glu Leu Lys Glu Arg Ser Gln Arg Pro
245 250 255
Val Arg Trp Met Leu Gln Asp Ile Gln Glu Val Leu Asn Arg Ser
260 265 270
Lys Ser Trp Ser Leu Gln Gln Pro Glu Pro Ile Ser Leu Glu Leu
275 280 285
Lys Thr Asp Cys Arg Val Leu Gly Leu Arg Glu Ile Leu Lys Thr
290 295 300
Tyr Ala Ala Asp Val Arg Leu Asp Pro Asp Thr Ala Tyr Ser Arg
305 310 315
Leu Ile Val Ser Glu Asp Arg Lys Arg Val His Tyr Gly Asp Thr
320 325 330
Asn Gln Lys Leu Pro Asp Asn Pro Glu Arg Phe Tyr Arg Tyr Asn
335 340 345
Ile Val Leu Gly Ser Gln Cys Ile Ser Ser Gly Arg His Tyr Trp
350 355 360
Glu Val Glu Val Gly Asp Arg Ser Glu Trp Gly Leu Gly Val Cys
365 370 375
Lys Gln Asn Val Asp Arg Lys Glu Val Val Tyr Leu Ser Pro His
380 385 390
Tyr Gly Phe Trp Val Ile Arg Leu Arg Lys Gly Asn Glu Tyr Arg
395 400 405
Ala Gly Thr Asp Glu Tyr Pro Ile Leu Ser Leu Pro Val Pro Pro
410 415 420
Arg Arg Val Gly Ile Phe Val Asp Tyr Glu Ala His Asp Ile Ser
425 430 435
Phe Tyr Asn Val Thr Asp Cys Gly Ser His Ile Phe Thr Phe Pro
440 445 450
Arg Tyr Pro Phe Pro Gly Arg Leu Leu Pro Tyr Phe Ser Pro Cys
455 460 465

Tyr Ser Ile Gly Thr Asn Asn Thr Ala Pro Leu Ala Ile Cys Ser
470 475 480
Leu Asp Gly Glu Asp
485

<210> 2
<211> 1404
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3125036CD1

<400> 2
Met Glu Ser Ser Ser Ser Asp Tyr Tyr Asn Lys Asp Asn Glu Glu
1 5 10 15
Glu Ser Leu Leu Ala Asn Val Ala Ser Leu Arg His Glu Leu Lys
20 25 30
Ile Thr Glu Trp Ser Leu Gln Ser Leu Gly Glu Glu Leu Ser Ser
35 40 45
Val Ser Pro Ser Glu Asn Ser Asp Tyr Ala Pro Asn Pro Ser Arg
50 55 60
Ser Glu Lys Leu Ile Leu Asp Val Gln Pro Ser His Pro Gly Leu
65 70 75
Leu Asn Tyr Ser Pro Tyr Glu Asn Val Cys Lys Ile Ser Gly Ser
80 85 90
Ser Thr Asp Phe Gln Lys Lys Pro Arg Asp Lys Met Phe Ser Ser
95 100 105
Ser Ala Pro Val Asp Gln Glu Ile Lys Ser Leu Arg Glu Lys Leu
110 115 120
Asn Lys Leu Arg Gln Gln Asn Ala Cys Leu Val Thr Gln Asn His
125 130 135
Ser Leu Met Thr Lys Phe Glu Ser Ile His Phe Glu Leu Thr Gln
140 145 150
Ser Arg Ala Lys Val Ser Met Leu Glu Ser Ala Gln Gln Gln Ala
155 160 165
Ala Ser Val Pro Ile Leu Glu Glu Gln Ile Ile Asn Leu Glu Ala
170 175 180
Glu Val Ser Ala Gln Asp Lys Val Leu Arg Glu Ala Glu Asn Lys
185 190 195
Leu Glu Gln Ser Gln Lys Met Val Ile Glu Lys Glu Gln Ser Leu
200 205 210
Gln Glu Ser Lys Glu Glu Cys Ile Lys Leu Lys Val Asp Leu Leu
215 220 225
Glu Gln Thr Lys Gln Gly Lys Arg Ala Glu Arg Gln Arg Asn Glu
230 235 240
Ala Leu Tyr Asn Ala Glu Glu Leu Ser Lys Ala Phe Gln Gln Tyr
245 250 255
Lys Lys Lys Val Ala Glu Lys Leu Glu Lys Val Lys Gly Ser Cys
260 265 270
Ala Asn Ser Val Phe Cys Ile Thr Val Tyr Ile Pro Thr Val Lys
275 280 285
Val Gln Ala Glu Glu Glu Ile Leu Glu Arg Asn Leu Thr Asn Cys
290 295 300
Glu Lys Glu Asn Lys Arg Leu Gln Glu Arg Cys Gly Leu Tyr Lys

305 310 315
Ser Glu Leu Glu Ile Leu Lys Glu Lys Leu Arg Gln Leu Lys Glu
320 325 330
Glu Asn Asn Asn Gly Lys Glu Lys Leu Arg Ile Met Ala Val Lys
335 340 345
Asn Ser Glu Val Met Ala Gln Leu Thr Glu Ser Arg Gln Ser Ile
350 355 360
Leu Lys Leu Glu Ser Glu Leu Glu Asn Lys Asp Glu Ile Leu Arg
365 370 375
Asp Lys Phe Ser Leu Met Asn Glu Asn Arg Glu Leu Lys Val Arg
380 385 390
Val Ala Ala Gln Asn Glu Arg Leu Asp Leu Cys Gln Gln Glu Ile
395 400 405
Glu Ser Ser Arg Val Glu Leu Arg Ser Leu Glu Lys Ile Ile Ser
410 415 420
Gln Leu Pro Leu Lys Arg Glu Leu Phe Gly Phe Lys Ser Tyr Leu
425 430 435
Ser Lys Tyr Gln Met Ser Ser Phe Ser Asn Lys Glu Asp Arg Cys
440 445 450
Ile Gly Cys Cys Glu Ala Asn Lys Leu Val Ile Ser Glu Leu Arg
455 460 465
Ile Lys Leu Ala Ile Lys Glu Ala Glu Ile Gln Lys Leu His Ala
470 475 480
Asn Leu Thr Ala Asn Gln Leu Ser Gln Ser Leu Ile Thr Cys Asn
485 490 495
Asp Ser Gln Glu Ser Ser Lys Leu Ser Ser Leu Glu Thr Glu Pro
500 505 510
Val Lys Leu Gly Gly His Gln Val Ala Glu Ser Val Lys Asp Gln
515 520 525
Asn Gln His Thr Met Asn Lys Gln Tyr Glu Lys Glu Arg Gln Arg
530 535 540
Leu Val Thr Gly Ile Glu Glu Leu Arg Thr Lys Leu Ile Gln Ile
545 550 555
Glu Ala Glu Asn Ser Asp Leu Lys Val Asn Met Ala His Arg Thr
560 565 570
Ser Gln Phe Gln Leu Ile Gln Glu Glu Leu Leu Glu Lys Ala Ser
575 580 585
Asn Ser Ser Lys Leu Glu Ser Glu Met Thr Lys Lys Cys Ser Gln
590 595 600
Leu Leu Thr Leu Glu Lys Gln Leu Glu Glu Lys Ile Val Ala Tyr
605 610 615
Ser Ser Ile Ala Ala Lys Asn Ala Glu Leu Glu Gln Glu Leu Met
620 625 630
Glu Lys Asn Glu Lys Ile Arg Ser Leu Glu Thr Asn Ile Asn Thr
635 640 645
Glu His Glu Lys Ile Cys Leu Ala Phe Glu Lys Ala Lys Lys Ile
650 655 660
His Leu Glu Gln His Lys Glu Met Glu Lys Gln Ile Glu Arg Val
665 670 675
Arg Gln Leu Asp Ser Ala Leu Glu Ile Cys Lys Glu Glu Leu Val
680 685 690
Leu His Leu Asn Gln Leu Glu Gly Asn Lys Glu Lys Phe Glu Lys
695 700 705
Gln Leu Lys Lys Lys Ser Glu Glu Val Tyr Cys Leu Gln Lys Glu
710 715 720
Leu Lys Ile Lys Asn His Ser Leu Gln Glu Thr Ser Glu Gln Asn

```
                725                 730                 735
Val Ile Leu Gln His Thr Leu Gln Gln Gln Gln Gln Met Leu Gln
                740                 745                 750
Gln Glu Thr Ile Arg Asn Gly Glu Leu Glu Asp Thr Gln Thr Lys
                755                 760                 765
Leu Glu Lys Gln Val Ser Lys Leu Glu Gln Glu Leu Gln Lys Gln
                770                 775                 780
Arg Glu Ser Ser Ala Glu Lys Leu Arg Lys Met Glu Glu Lys Cys
                785                 790                 795
Glu Ser Ala Ala His Glu Ala Asp Leu Lys Arg Gln Lys Val Ile
                800                 805                 810
Glu Leu Thr Gly Thr Ala Arg Gln Val Lys Ile Glu Met Asp Gln
                815                 820                 825
Tyr Lys Glu Glu Leu Ser Lys Met Glu Lys Glu Ile Met His Leu
                830                 835                 840
Lys Arg Asp Gly Glu Asn Lys Ala Met His Leu Ser Gln Leu Asp
                845                 850                 855
Met Ile Leu Asp Gln Thr Lys Thr Glu Leu Glu Lys Lys Thr Asn
                860                 865                 870
Ala Val Lys Glu Leu Glu Lys Leu Gln His Ser Thr Glu Thr Glu
                875                 880                 885
Leu Thr Glu Ala Leu Gln Lys Arg Glu Val Leu Glu Thr Glu Leu
                890                 895                 900
Gln Asn Ala His Gly Glu Leu Lys Ser Thr Leu Arg Gln Leu Gln
                905                 910                 915
Glu Leu Arg Asp Val Leu Gln Lys Ala Gln Leu Ser Leu Glu Glu
                920                 925                 930
Lys Tyr Thr Thr Ile Lys Asp Leu Thr Ala Glu Leu Arg Glu Cys
                935                 940                 945
Lys Met Glu Ile Glu Asp Lys Lys Gln Glu Leu Leu Glu Met Asp
                950                 955                 960
Gln Ala Leu Lys Glu Arg Asn Trp Glu Leu Lys Gln Arg Ala Ala
                965                 970                 975
Gln Val Thr His Leu Asp Met Thr Ile Arg Glu His Arg Gly Glu
                980                 985                 990
Met Glu Gln Lys Ile Ile Lys Leu Glu Gly Thr Leu Glu Lys Ser
                995                 1000                1005
Glu Leu Glu Leu Lys Glu Cys Asn Lys Gln Ile Glu Ser Leu Asn
                1010                1015                1020
Asp Lys Leu Gln Asn Ala Lys Glu Gln Val Arg Glu Lys Glu Phe
                1025                1030                1035
Ile Met Leu Gln Asn Glu Gln Glu Ile Ser Gln Leu Lys Lys Glu
                1040                1045                1050
Ile Glu Arg Thr Gln Gln Arg Met Lys Glu Met Glu Ser Val Met
                1055                1060                1065
Lys Glu Gln Glu Gln Tyr Ile Ala Thr Gln Tyr Lys Glu Ala Ile
                1070                1075                1080
Asp Leu Gly Gln Glu Leu Arg Leu Thr Arg Glu Gln Val Gln Asn
                1085                1090                1095
Ser His Thr Glu Leu Ala Glu Ala Arg His Gln Gln Val Gln Ala
                1100                1105                1110
Gln Arg Glu Ile Glu Arg Leu Ser Ser Glu Leu Glu Asp Met Lys
                1115                1120                1125
Gln Leu Ser Lys Glu Lys Asp Ala His Gly Asn His Leu Ala Glu
                1130                1135                1140
Glu Leu Gly Ala Ser Lys Val Arg Glu Ala His Leu Glu Ala Arg
```

1145 1150 1155
Met Gln Ala Glu Ile Lys Lys Leu Ser Ala Glu Val Glu Ser Leu
1160 1165 1170
Lys Glu Ala Tyr His Met Glu Met Ile Ser His Gln Glu Asn His
1175 1180 1185
Ala Lys Trp Lys Ile Ser Ala Asp Ser Gln Lys Ser Ser Val Gln
1190 1195 1200
Gln Leu Asn Glu Gln Leu Glu Lys Ala Lys Leu Glu Leu Glu Glu
1205 1210 1215
Ala Gln Asp Thr Val Ser Asn Leu His Gln Gln Val Gln Asp Arg
1220 1225 1230
Asn Glu Val Ile Glu Ala Ala Asn Glu Ala Leu Leu Thr Lys Glu
1235 1240 1245
Ser Glu Leu Thr Arg Leu Gln Ala Lys Ile Ser Gly His Glu Lys
1250 1255 1260
Ala Glu Asp Ile Lys Phe Leu Pro Ala Pro Phe Thr Ser Pro Thr
1265 1270 1275
Glu Ile Met Pro Asp Val Gln Asp Pro Lys Phe Ala Lys Cys Phe
1280 1285 1290
His Thr Ser Phe Ser Lys Cys Thr Lys Leu Arg Arg Ser Ile Ser
1295 1300 1305
Ala Ser Asp Leu Thr Phe Lys Ile His Gly Asp Glu Asp Leu Ser
1310 1315 1320
Glu Glu Leu Leu Gln Asp Leu Lys Lys Met Gln Leu Glu Gln Pro
1325 1330 1335
Ser Thr Leu Glu Glu Ser His Lys Asn Leu Thr Tyr Thr Gln Pro
1340 1345 1350
Asp Ser Phe Lys Pro Leu Thr Tyr Asn Leu Glu Ala Asp Ser Ser
1355 1360 1365
Glu Asn Asn Asp Phe Asn Thr Leu Ser Gly Met Leu Arg Tyr Ile
1370 1375 1380
Asn Lys Glu Val Arg Leu Leu Lys Lys Ser Ser Met Gln Thr Gly
1385 1390 1395
Ala Gly Leu Asn Gln Gly Glu Asn Val
1400

<210> 3
<211> 1096
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1758089CD1

<400> 3
Met Gly Ser Glu Asp His Gly Ala Gln Asn Pro Ser Cys Lys Ile
1 5 10 15
Met Thr Phe Arg Pro Thr Met Glu Glu Phe Lys Asp Phe Asn Lys
20 25 30
Tyr Val Ala Tyr Ile Glu Ser Gln Gly Ala His Arg Ala Gly Leu
35 40 45
Ala Lys Ile Ile Pro Pro Lys Glu Trp Lys Pro Arg Gln Thr Tyr
50 55 60
Asp Asp Ile Asp Asp Val Val Ile Pro Ala Pro Ile Gln Gln Val
65 70 75

Val Thr Gly Gln Ser Gly Leu Phe Thr Gln Tyr Asn Ile Gln Lys
80 85 90
Lys Ala Met Thr Val Gly Glu Tyr Arg Arg Leu Ala Asn Ser Glu
95 100 105
Lys Tyr Cys Thr Pro Arg His Gln Asp Phe Asp Asp Leu Glu Arg
110 115 120
Lys Tyr Trp Lys Asn Leu Thr Phe Val Ser Pro Ile Tyr Gly Ala
125 130 135
Asp Ile Ser Gly Ser Leu Tyr Asp Asp Asp Val Ala Gln Trp Asn
140 145 150
Ile Gly Ser Leu Arg Thr Ile Leu Asp Met Val Glu Arg Glu Cys
155 160 165
Gly Thr Ile Ile Glu Gly Val Asn Thr Pro Tyr Leu Tyr Phe Gly
170 175 180
Met Trp Lys Thr Thr Phe Ala Trp His Thr Glu Asp Met Asp Leu
185 190 195
Tyr Ser Ile Asn Tyr Leu His Phe Gly Glu Pro Lys Ser Trp Tyr
200 205 210
Ala Ile Pro Pro Glu His Gly Lys Arg Leu Glu Arg Leu Ala Ile
215 220 225
Gly Phe Phe Pro Gly Ser Ser Gln Gly Cys Asp Ala Phe Leu Arg
230 235 240
His Lys Met Thr Leu Ile Ser Pro Ile Ile Leu Lys Lys Tyr Gly
245 250 255
Ile Pro Phe Ser Arg Ile Thr Gln Glu Ala Gly Glu Phe Met Ile
260 265 270
Thr Phe Pro Tyr Gly Tyr His Ala Gly Phe Asn His Gly Phe Asn
275 280 285
Cys Ala Glu Ser Thr Asn Phe Ala Thr Leu Arg Trp Ile Asp Tyr
290 295 300
Gly Lys Val Ala Thr Gln Cys Thr Cys Arg Lys Asp Met Val Lys
305 310 315
Ile Ser Met Asp Val Phe Val Arg Ile Leu Gln Pro Glu Arg Tyr
320 325 330
Glu Leu Trp Lys Gln Gly Lys Asp Leu Thr Val Leu Asp His Thr
335 340 345
Arg Pro Thr Ala Leu Thr Ser Pro Glu Leu Ser Ser Trp Ser Ala
350 355 360
Ser Arg Ala Ser Leu Lys Ala Lys Leu Leu Arg Arg Ser His Arg
365 370 375
Lys Arg Ser Gln Pro Lys Lys Pro Lys Pro Glu Asp Pro Lys Phe
380 385 390
Pro Gly Glu Gly Thr Ala Gly Ala Ala Leu Leu Glu Glu Ala Gly
395 400 405
Gly Ser Val Lys Glu Glu Ala Gly Pro Glu Val Asp Pro Glu Glu
410 415 420
Glu Glu Glu Glu Pro Gln Pro Leu Pro His Gly Arg Glu Ala Glu
425 430 435
Gly Ala Glu Glu Asp Gly Arg Gly Lys Leu Arg Pro Thr Lys Ala
440 445 450
Lys Ser Glu Arg Lys Lys Lys Ser Phe Gly Leu Leu Pro Pro Gln
455 460 465
Leu Pro Pro Pro Pro Ala His Phe Pro Ser Glu Glu Ala Leu Trp
470 475 480
Leu Pro Ser Pro Leu Glu Pro Pro Val Leu Gly Pro Gly Pro Ala
485 490 495

Ala Met Glu Glu Ser Pro Leu Pro Ala Pro Leu Asn Val Val Pro
500 505 510
Pro Glu Val Pro Ser Glu Glu Leu Glu Ala Lys Pro Arg Pro Ile
515 520 525
Ile Pro Met Leu Tyr Val Val Pro Arg Pro Gly Lys Ala Ala Phe
530 535 540
Asn Gln Glu His Val Ser Cys Gln Gln Ala Phe Glu His Phe Ala
545 550 555
Gln Lys Gly Pro Thr Trp Lys Glu Pro Val Ser Pro Met Glu Leu
560 565 570
Thr Gly Pro Glu Asp Gly Ala Ala Ser Ser Gly Ala Gly Arg Met
575 580 585
Glu Thr Lys Ala Arg Ala Gly Glu Gly Gln Ala Pro Ser Thr Phe
590 595 600
Ser Lys Leu Lys Met Glu Ile Lys Lys Ser Arg Arg His Pro Leu
605 610 615
Gly Arg Pro Pro Thr Arg Ser Pro Leu Ser Val Val Lys Gln Glu
620 625 630
Ala Ser Ser Asp Glu Glu Ala Ser Pro Phe Ser Gly Glu Glu Asp
635 640 645
Val Ser Asp Pro Asp Ala Leu Arg Pro Leu Leu Ser Leu Gln Trp
650 655 660
Lys Asn Arg Ala Ala Ser Phe Gln Ala Glu Arg Lys Phe Asn Ala
665 670 675
Ala Ala Ala Arg Thr Glu Pro Tyr Cys Ala Ile Cys Thr Leu Phe
680 685 690
Tyr Pro Tyr Cys Gln Ala Leu Gln Thr Glu Lys Glu Ala Pro Ile
695 700 705
Ala Ser Leu Gly Glu Gly Cys Pro Ala Thr Leu Pro Ser Lys Ser
710 715 720
Arg Gln Lys Thr Arg Pro Leu Ile Pro Glu Met Cys Phe Thr Ser
725 730 735
Gly Gly Glu Asn Thr Glu Pro Leu Pro Ala Asn Ser Tyr Ile Gly
740 745 750
Asp Asp Gly Thr Ser Pro Leu Ile Ala Cys Gly Lys Cys Cys Leu
755 760 765
Gln Val His Ala Ser Cys Tyr Gly Ile Arg Pro Glu Leu Val Asn
770 775 780
Glu Gly Trp Thr Cys Ser Arg Cys Ala Ala His Ala Trp Thr Ala
785 790 795
Glu Cys Cys Leu Cys Asn Leu Arg Gly Gly Ala Leu Gln Met Thr
800 805 810
Thr Asp Arg Arg Trp Ile His Val Ile Cys Ala Ile Ala Val Pro
815 820 825
Glu Ala Arg Phe Leu Asn Val Ile Glu Arg His Pro Val Asp Ile
830 835 840
Ser Ala Ile Pro Glu Gln Arg Trp Lys Leu Lys Cys Val Tyr Cys
845 850 855
Arg Lys Arg Met Lys Lys Val Ser Gly Ala Cys Ile Gln Cys Ser
860 865 870
Tyr Glu His Cys Ser Thr Ser Phe His Val Thr Cys Ala His Ala
875 880 885
Ala Gly Val Leu Met Glu Pro Asp Asp Trp Pro Tyr Val Val Ser
890 895 900
Ile Thr Cys Leu Lys His Lys Ser Gly Gly His Ala Val Gln Leu
905 910 915

Leu Arg Ala Val Ser Leu Gly Gln Val Val Ile Thr Lys Asn Arg
920 925 930
Asn Gly Leu Tyr Tyr Arg Cys Arg Val Ile Gly Ala Ala Ser Gln
935 940 945
Thr Cys Tyr Glu Val Asn Phe Asp Asp Gly Ser Tyr Ser Asp Asn
950 955 960
Leu Tyr Pro Glu Ser Ile Thr Ser Arg Asp Cys Val Gln Leu Gly
965 970 975
Pro Pro Ser Glu Gly Glu Leu Val Glu Leu Arg Trp Thr Asp Gly
980 985 990
Asn Leu Tyr Lys Ala Lys Phe Ile Ser Ser Val Thr Ser His Ile
995 1000 1005
Tyr Gln Val Glu Phe Glu Asp Gly Ser Gln Leu Thr Val Lys Arg
1010 1015 1020
Gly Asp Ile Phe Thr Leu Glu Glu Glu Leu Pro Lys Arg Val Arg
1025 1030 1035
Ser Arg Leu Ser Leu Ser Thr Gly Ala Pro Gln Glu Pro Ala Phe
1040 1045 1050
Ser Gly Glu Glu Ala Lys Ala Ala Lys Arg Pro Arg Val Gly Thr
1055 1060 1065
Pro Leu Ala Thr Glu Asp Ser Gly Arg Ser Gln Asp Tyr Val Ala
1070 1075 1080
Phe Val Glu Ser Leu Leu Gln Val Gln Gly Arg Pro Gly Ala Pro
1085 1090 1095
Phe

<210> 4
<211> 167
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3533891CD1

<400> 4
Met Tyr Met Gly Met Met Cys Thr Ala Lys Lys Cys Gly Ile Arg
1 5 10 15
Phe Gln Pro Pro Ala Ile Ile Leu Ile Tyr Glu Ser Glu Ile Lys
20 25 30
Gly Lys Ile Arg Gln Arg Ile Met Pro Val Arg Asn Phe Ser Lys
35 40 45
Phe Ser Asp Cys Thr Arg Ala Ala Glu Gln Leu Lys Asn Asn Pro
50 55 60
Arg His Lys Ser Tyr Leu Glu Gln Val Ser Leu Arg Gln Leu Glu
65 70 75
Lys Leu Phe Ser Phe Leu Arg Gly Tyr Leu Ser Gly Gln Ser Leu
80 85 90
Ala Glu Thr Met Glu Gln Ile Gln Arg Glu Thr Thr Ile Asp Pro
95 100 105
Glu Glu Asp Leu Asn Lys Leu Asp Asp Lys Glu Leu Ala Lys Arg
110 115 120
Lys Ser Ile Met Asp Glu Leu Phe Glu Lys Asn Gln Lys Lys Lys
125 130 135
Asp Asp Pro Asn Phe Val Tyr Asp Ile Glu Val Glu Phe Pro Gln

140 145 150
Asp Asp Gln Leu Gln Ser Cys Gly Trp Asp Thr Glu Ser Ala Asp
155 160 165
Glu Phe

<210> 5
<211> 1523
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1510943CD1

<400> 5
Met Thr Ser Val Trp Lys Arg Leu Gln Arg Val Gly Lys Arg Ala
1 5 10 15
Ala Lys Phe Gln Phe Val Ala Cys Tyr His Glu Leu Val Leu Glu
20 25 30
Cys Thr Lys Lys Trp Gln Pro Asp Lys Leu Val Val Val Trp Thr
35 40 45
Arg Arg Asn Arg Arg Ile Cys Ser Lys Ala His Ser Trp Gln Pro
50 55 60
Gly Ile Gln Asn Pro Tyr Arg Gly Thr Val Val Trp Met Val Pro
65 70 75
Glu Asn Val Asp Ile Ser Val Thr Leu Tyr Arg Asp Pro His Val
80 85 90
Asp Gln Tyr Glu Ala Lys Glu Trp Thr Phe Ile Ile Glu Asn Glu
95 100 105
Ser Lys Gly Gln Arg Lys Val Leu Ala Thr Ala Glu Val Asp Leu
110 115 120
Ala Arg His Ala Gly Pro Val Pro Val Gln Val Pro Leu Arg Leu
125 130 135
Arg Leu Lys Pro Lys Ser Val Lys Val Val Gln Ala Glu Leu Ser
140 145 150
Leu Thr Leu Ser Gly Val Leu Leu Arg Glu Gly Arg Ala Thr Asp
155 160 165
Asp Asp Met Gln Ser Leu Ala Ser Leu Met Ser Val Lys Pro Ser
170 175 180
Asp Val Gly Asn Leu Asp Asp Phe Ala Glu Ser Asp Glu Asp Glu
185 190 195
Ala His Gly Pro Gly Ala Pro Glu Ala Arg Ala Arg Val Pro Gln
200 205 210
Pro Asp Pro Ser Arg Glu Leu Lys Thr Leu Cys Glu Glu Glu Glu
215 220 225
Glu Gly Gln Gly Arg Pro Gln Gln Ala Val Ala Ser Pro Ser Asn
230 235 240
Ala Glu Asp Thr Ser Pro Ala Pro Val Ser Ala Pro Ala Pro Pro
245 250 255
Ala Arg Thr Ser Arg Gly Gln Gly Ser Glu Arg Ala Asn Glu Ala
260 265 270
Gly Gly Gln Val Gly Pro Glu Ala Pro Arg Pro Pro Glu Thr Ser
275 280 285
Pro Glu Met Arg Ser Ser Arg Gln Pro Ala Gln Asp Thr Ala Pro
290 295 300

Thr Pro Ala Pro Arg Leu Arg Lys Gly Ser Asp Ala Leu Arg Pro
305 310 315
Pro Val Pro Gln Gly Glu Asp Glu Val Pro Lys Ala Ser Gly Ala
320 325 330
Pro Pro Ala Gly Leu Gly Ser Ala Arg Glu Thr Gln Ala Gln Ala
335 340 345
Cys Pro Gln Glu Gly Thr Glu Ala His Gly Ala Arg Leu Gly Pro
350 355 360
Ser Ile Glu Asp Lys Gly Ser Gly Asp Pro Phe Gly Arg Gln Arg
365 370 375
Leu Lys Ala Glu Glu Met Asp Thr Glu Asp Arg Pro Glu Ala Ser
380 385 390
Gly Val Asp Thr Glu Pro Arg Ser Gly Gly Arg Glu Ala Asn Thr
395 400 405
Lys Arg Ser Gly Val Arg Ala Gly Glu Ala Glu Glu Ser Ser Ala
410 415 420
Val Cys Gln Val Asp Ala Glu Gln Arg Ser Lys Val Arg His Val
425 430 435
Asp Thr Lys Gly Pro Glu Ala Thr Gly Val Met Pro Glu Ala Arg
440 445 450
Cys Arg Gly Thr Pro Glu Ala Pro Pro Arg Gly Ser Gln Gly Arg
455 460 465
Leu Gly Val Arg Thr Arg Asp Glu Ala Pro Ser Gly Leu Ser Leu
470 475 480
Pro Pro Ala Glu Pro Ala Gly His Ser Gly Gln Leu Gly Asp Leu
485 490 495
Glu Gly Ala Arg Ala Ala Ala Gly Gln Glu Arg Glu Gly Ala Glu
500 505 510
Val Arg Gly Gly Ala Pro Gly Ile Glu Gly Thr Gly Leu Glu Gln
515 520 525
Gly Pro Ser Val Gly Ala Ile Ser Thr Arg Pro Gln Val Ser Ser
530 535 540
Trp Gln Gly Ala Leu Leu Ser Thr Ala Gln Gly Ala Ile Ser Arg
545 550 555
Gly Leu Gly Gly Trp Glu Ala Glu Ala Gly Gly Ser Gly Val Leu
560 565 570
Glu Thr Glu Thr Glu Val Val Gly Leu Glu Val Leu Gly Thr Gln
575 580 585
Glu Lys Glu Val Glu Gly Ser Gly Phe Pro Glu Thr Arg Thr Leu
590 595 600
Glu Ile Glu Ile Leu Gly Ala Leu Glu Lys Glu Ala Ala Arg Ser
605 610 615
Arg Val Leu Glu Ser Glu Val Ala Gly Thr Ala Gln Cys Glu Gly
620 625 630
Leu Glu Thr Gln Glu Thr Glu Val Gly Val Ile Glu Thr Pro Gly
635 640 645
Thr Glu Thr Glu Val Leu Gly Thr Gln Lys Thr Glu Ala Gly Gly
650 655 660
Ser Gly Val Leu Gln Thr Arg Thr Thr Ile Ala Glu Thr Glu Val
665 670 675
Leu Val Thr Gln Glu Ile Ser Gly Asp Leu Gly Pro Leu Lys Ile
680 685 690
Glu Asp Thr Ile Gln Ser Glu Met Leu Gly Thr Gln Glu Thr Glu
695 700 705
Val Glu Ala Ser Arg Val Pro Glu Ser Glu Ala Glu Gly Thr Glu
710 715 720

```
Ala Lys Ile Leu Gly Thr Gln Glu Ile Thr Ala Arg Asp Ser Gly
                725                 730                 735
Val Arg Glu Ile Glu Ala Glu Ile Ala Glu Ser Asp Ile Leu Val
                740                 745                 750
Ala Gln Glu Ile Glu Val Gly Leu Leu Gly Val Leu Gly Ile Glu
                755                 760                 765
Thr Gly Ala Ala Glu Gly Ala Ile Leu Gly Thr Gln Glu Ile Ala
                770                 775                 780
Ser Arg Asp Ser Gly Val Pro Gly Leu Glu Ala Asp Thr Thr Gly
                785                 790                 795
Ile Gln Val Lys Glu Val Gly Gly Ser Glu Val Pro Glu Ile Ala
                800                 805                 810
Thr Gly Thr Ala Glu Thr Glu Ile Leu Gly Thr Gln Glu Ile Ala
                815                 820                 825
Ser Arg Ser Ser Gly Val Pro Gly Leu Glu Ser Glu Val Ala Gly
                830                 835                 840
Ala Gln Glu Thr Glu Val Gly Gly Ser Gly Ile Ser Gly Pro Glu
                845                 850                 855
Ala Gly Met Ala Glu Ala Arg Val Leu Met Thr Arg Lys Thr Glu
                860                 865                 870
Ile Ile Val Pro Glu Ala Glu Lys Glu Glu Ala Gln Thr Ser Gly
                875                 880                 885
Val Gln Glu Ala Glu Thr Arg Val Gly Ser Ala Leu Lys Tyr Glu
                890                 895                 900
Ala Leu Arg Ala Pro Val Thr Gln Pro Arg Val Leu Gly Ser Gln
                905                 910                 915
Glu Ala Lys Ala Glu Ile Ser Gly Val Gln Gly Ser Glu Thr Gln
                920                 925                 930
Val Leu Arg Val Gln Glu Ala Glu Ala Gly Val Trp Gly Met Ser
                935                 940                 945
Glu Gly Lys Ser Gly Ala Trp Gly Ala Gln Glu Ala Glu Met Lys
                950                 955                 960
Val Leu Glu Ser Pro Glu Asn Lys Ser Gly Thr Phe Lys Ala Gln
                965                 970                 975
Glu Ala Glu Ala Gly Val Leu Gly Asn Glu Lys Gly Lys Glu Ala
                980                 985                 990
Glu Gly Ser Leu Thr Glu Ala Ser Leu Pro Glu Ala Gln Val Ala
                995                 1000                1005
Ser Gly Ala Gly Ala Gly Ala Pro Arg Ala Ser Ser Pro Glu Lys
                1010                1015                1020
Ala Glu Glu Asp Arg Arg Leu Pro Gly Ser Gln Ala Pro Pro Ala
                1025                1030                1035
Leu Val Ser Ser Ser Gln Ser Leu Leu Glu Trp Cys Gln Glu Val
                1040                1045                1050
Thr Thr Gly Tyr Arg Gly Val Arg Ile Thr Asn Phe Thr Thr Ser
                1055                1060                1065
Trp Arg Asn Gly Leu Ala Phe Cys Ala Ile Leu His Arg Phe Tyr
                1070                1075                1080
Pro Asp Lys Ile Asp Tyr Ala Ser Leu Asp Pro Leu Asn Ile Lys
                1085                1090                1095
Gln Asn Asn Lys Gln Ala Phe Asp Gly Phe Ala Ala Leu Gly Val
                1100                1105                1110
Ser Arg Leu Leu Glu Pro Ala Asp Met Val Leu Leu Ser Val Pro
                1115                1120                1125
Asp Lys Leu Ile Val Met Thr Tyr Leu Cys Gln Ile Arg Ala Phe
                1130                1135                1140
```

Cys Thr Gly Gln Glu Leu Gln Leu Val Gln Leu Glu Gly Gly Gly
1145 1150 1155
Gly Ala Gly Thr Tyr Arg Val Gly Ser Ala Gln Pro Ser Pro Pro
1160 1165 1170
Asp Asp Leu Asp Ala Gly Gly Leu Ala Gln Arg Leu Arg Gly His
1175 1180 1185
Gly Ala Glu Gly Pro Gln Glu Pro Lys Glu Ala Ala Asp Arg Ala
1190 1195 1200
Asp Gly Ala Ala Pro Gly Val Ala Ser Arg Asn Ala Val Ala Gly
1205 1210 1215
Arg Ala Ser Lys Asp Gly Gly Ala Glu Ala Pro Arg Glu Ser Arg
1220 1225 1230
Pro Ala Glu Val Pro Ala Glu Gly Leu Val Asn Gly Ala Gly Ala
1235 1240 1245
Pro Gly Gly Gly Gly Val Arg Leu Arg Arg Pro Ser Val Asn Gly
1250 1255 1260
Glu Pro Gly Ser Val Pro Pro Pro Arg Ala His Gly Ser Phe Ser
1265 1270 1275
His Val Arg Asp Ala Asp Leu Leu Lys Lys Arg Arg Ser Arg Leu
1280 1285 1290
Arg Asn Ser Ser Ser Phe Ser Met Asp Asp Pro Asp Ala Gly Ala
1295 1300 1305
Met Gly Ala Ala Ala Ala Glu Gly Gln Ala Pro Asp Pro Ser Pro
1310 1315 1320
Ala Pro Gly Pro Pro Thr Ala Ala Asp Ser Gln Gln Pro Pro Gly
1325 1330 1335
Gly Ser Ser Pro Ser Glu Glu Pro Pro Pro Ser Pro Gly Glu Glu
1340 1345 1350
Ala Gly Leu Gln Arg Phe Gln Asp Thr Ser Gln Tyr Val Cys Ala
1355 1360 1365
Glu Leu Gln Ala Leu Glu Gln Glu Gln Arg Gln Ile Asp Gly Arg
1370 1375 1380
Ala Ala Glu Val Glu Met Gln Leu Arg Ser Leu Met Glu Ser Gly
1385 1390 1395
Ala Asn Lys Leu Gln Glu Glu Val Leu Ile Gln Glu Trp Phe Thr
1400 1405 1410
Leu Val Asn Lys Lys Asn Ala Leu Ile Arg Arg Gln Asp Gln Leu
1415 1420 1425
Gln Leu Leu Met Glu Glu Gln Asp Leu Glu Arg Arg Phe Glu Leu
1430 1435 1440
Leu Ser Arg Glu Leu Arg Ala Met Leu Ala Ile Glu Asp Trp Gln
1445 1450 1455
Lys Thr Ser Ala Gln Gln His Arg Glu Gln Leu Leu Leu Glu Glu
1460 1465 1470
Leu Val Ser Leu Val Asn Gln Arg Asp Glu Leu Val Arg Asp Leu
1475 1480 1485
Asp His Lys Glu Arg Ile Ala Leu Glu Glu Asp Glu Arg Leu Glu
1490 1495 1500
Arg Gly Leu Glu Gln Arg Arg Arg Lys Leu Ser Arg Gln Leu Ser
1505 1510 1515
Arg Arg Glu Arg Cys Val Leu Ser
1520

<210> 6
<211> 273
<212> PRT

<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2119377CD1

<400> 6
Met Gly Gln Lys Leu Ser Gly Ser Leu Lys Ser Val Glu Val Arg
1 5 10 15
Glu Pro Ala Leu Arg Pro Ala Lys Arg Glu Leu Arg Gly Ala Glu
20 25 30
Pro Gly Arg Pro Ala Arg Leu Asp Gln Leu Leu Asp Met Pro Ala
35 40 45
Ala Gly Leu Ala Val Gln Leu Arg His Ala Trp Asn Pro Glu Asp
50 55 60
Arg Ser Leu Asn Val Phe Val Lys Asp Asp Asp Arg Leu Thr Phe
65 70 75
His Arg His Pro Val Ala Gln Ser Thr Asp Gly Ile Arg Gly Lys
80 85 90
Val Gly His Ala Arg Gly Leu His Ala Trp Gln Ile Asn Trp Pro
95 100 105
Ala Arg Gln Arg Gly Thr His Ala Val Val Gly Val Ala Thr Ala
110 115 120
Arg Ala Pro Leu His Ser Val Gly Tyr Thr Ala Leu Val Gly Ser
125 130 135
Asp Ala Glu Ser Trp Gly Trp Asp Leu Gly Arg Ser Arg Leu Tyr
140 145 150
His Asp Gly Lys Asn Gln Pro Gly Val Ala Tyr Pro Ala Phe Leu
155 160 165
Gly Pro Asp Glu Ala Phe Ala Leu Pro Asp Ser Leu Leu Val Val
170 175 180
Leu Asp Met Asp Glu Gly Thr Leu Ser Phe Ile Val Asp Gly Gln
185 190 195
Tyr Leu Gly Val Ala Phe Arg Gly Leu Lys Gly Lys Lys Leu Tyr
200 205 210
Pro Val Val Ser Ala Val Trp Gly His Cys Glu Val Thr Met Arg
215 220 225
Tyr Ile Asn Gly Leu Asp Pro Glu Pro Leu Pro Leu Met Asp Leu
230 235 240
Cys Arg Arg Ser Ile Arg Ser Ala Leu Gly Arg Gln Arg Leu Gln
245 250 255
Asp Ile Ser Ser Leu Pro Leu Pro Gln Ser Leu Lys Asn Tyr Leu
260 265 270
Gln Tyr Gln

<210> 7
<211> 341
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3176058CD1

<400> 7

Met Asp Gly Leu Leu Asn Pro Arg Glu Ser Ser Lys Phe Ile Ala
1 5 10 15
Glu Asn Ser Arg Asp Val Phe Ile Asp Ser Gly Gly Val Arg Arg
20 25 30
Val Ala Glu Leu Leu Leu Ala Lys Ala Ala Gly Pro Glu Leu Arg
35 40 45
Val Glu Gly Trp Lys Ala Leu His Glu Leu Asn Pro Arg Ala Ala
50 55 60
Asp Glu Ala Ala Val Asn Trp Val Phe Val Thr Asp Thr Leu Asn
65 70 75
Phe Ser Phe Trp Ser Glu Gln Asp Glu His Lys Cys Val Val Arg
80 85 90
Tyr Arg Gly Lys Thr Tyr Ser Gly Tyr Trp Ser Leu Cys Ala Ala
95 100 105
Val Asn Arg Ala Leu Asp Glu Gly Ile Pro Ile Thr Ser Ala Ser
110 115 120
Tyr Tyr Ala Thr Val Thr Leu Asp Gln Val Arg Asn Ile Leu Arg
125 130 135
Ser Asp Thr Asp Val Ser Met Pro Leu Val Glu Glu Arg His Arg
140 145 150
Ile Leu Asn Glu Thr Gly Lys Ile Leu Leu Glu Lys Phe Gly Gly
155 160 165
Ser Phe Leu Asn Cys Val Arg Glu Ser Glu Asn Ser Ala Gln Lys
170 175 180
Leu Met His Leu Val Val Glu Ser Phe Pro Ser Tyr Arg Asp Val
185 190 195
Thr Leu Phe Glu Gly Lys Arg Val Ser Phe Tyr Lys Arg Ala Gln
200 205 210
Ile Leu Val Ala Asp Thr Trp Ser Val Leu Glu Gly Lys Gly Asp
215 220 225
Gly Cys Phe Lys Asp Ile Ser Ser Ile Thr Met Phe Ala Asp Tyr
230 235 240
Arg Leu Pro Gln Val Leu Ala His Leu Gly Ala Leu Lys Tyr Ser
245 250 255
Asp Asp Leu Leu Lys Lys Leu Leu Lys Gly Glu Met Leu Ser Tyr
260 265 270
Gly Asp Arg Gln Glu Val Glu Ile Arg Gly Cys Ser Leu Trp Cys
275 280 285
Val Glu Leu Ile Arg Asp Cys Leu Leu Glu Leu Ile Glu Gln Lys
290 295 300
Gly Glu Lys Pro Asn Gly Glu Ile Asn Ser Ile Leu Leu Asp Tyr
305 310 315
Tyr Leu Trp Asp Tyr Ala His Asp His Arg Glu Asp Met Lys Gly
320 325 330
Ile Pro Phe His Arg Ile Arg Cys Ile Tyr Tyr
335 340

<210> 8
<211> 341
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2299818CD1

<400> 8
Met Asn Phe Lys Leu Gly Asn Phe Ser Tyr Gln Lys Asn Pro Leu
1   5   10   15
Lys Leu Gly Glu Leu Gln Gly Asn His Phe Thr Val Val Leu Arg
                20   25   30
Asn Ile Thr Gly Thr Asp Asp Gln Val Gln Gln Ala Met Asn Ser
                35   40   45
Leu Lys Glu Ile Gly Phe Ile Asn Tyr Tyr Gly Met Gln Arg Phe
                50   55   60
Gly Thr Thr Ala Val Pro Thr Tyr Gln Val Gly Arg Ala Ile Leu
                65   70   75
Gln Asn Ser Trp Thr Glu Val Met Asp Leu Ile Leu Lys Pro Arg
                80   85   90
Ser Gly Ala Glu Lys Gly Tyr Leu Val Lys Cys Arg Glu Glu Trp
                95   100   105
Ala Lys Thr Lys Asp Pro Thr Ala Ala Leu Arg Lys Leu Pro Val
                110   115   120
Lys Arg Cys Val Glu Gly Gln Leu Leu Arg Gly Leu Ser Lys Tyr
                125   130   135
Gly Met Lys Asn Ile Val Ser Ala Phe Gly Ile Ile Pro Arg Asn
                140   145   150
Asn Arg Leu Met Tyr Ile His Ser Tyr Gln Ser Tyr Val Trp Asn
                155   160   165
Asn Met Val Ser Lys Arg Ile Glu Asp Tyr Gly Leu Lys Pro Val
                170   175   180
Pro Gly Asp Leu Val Leu Lys Gly Ala Thr Ala Thr Tyr Ile Glu
                185   190   195
Glu Asp Asp Val Asn Asn Tyr Ser Ile His Asp Val Val Met Pro
                200   205   210
Leu Pro Gly Phe Asp Val Ile Tyr Pro Lys His Lys Ile Gln Glu
                215   220   225
Ala Tyr Arg Glu Met Leu Thr Ala Asp Asn Leu Asp Ile Asp Asn
                230   235   240
Met Arg His Lys Ile Arg Asp Tyr Ser Leu Ser Gly Ala Tyr Arg
                245   250   255
Lys Ile Ile Ile Arg Pro Gln Asn Val Ser Trp Glu Val Val Ala
                260   265   270
Tyr Asp Asp Pro Lys Ile Pro Leu Phe Asn Thr Asp Val Asp Asn
                275   280   285
Leu Glu Gly Lys Thr Pro Pro Val Phe Ala Ser Glu Gly Lys Tyr
                290   295   300
Arg Ala Leu Lys Met Asp Phe Ser Leu Pro Pro Ser Thr Tyr Ala
                305   310   315
Thr Met Ala Ile Arg Glu Val Leu Lys Met Asp Thr Ser Ile Lys
                320   325   330
Asn Gln Thr Gln Leu Asn Thr Thr Trp Leu Arg
                335   340

<210> 9
<211> 1185
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2729451CD1

<400> 9
Met Glu Pro Asn Ser Leu Gln Trp Val Gly Ser Pro Cys Gly Leu
1 5 10 15
His Gly Pro Tyr Ile Phe Tyr Lys Ala Phe Gln Phe His Leu Glu
20 25 30
Gly Lys Pro Arg Ile Leu Ser Leu Gly Asp Phe Phe Phe Val Arg
35 40 45
Cys Thr Pro Lys Asp Pro Ile Cys Ile Ala Glu Leu Gln Leu Leu
50 55 60
Trp Glu Glu Arg Thr Ser Arg Gln Leu Leu Ser Ser Ser Lys Leu
65 70 75
Tyr Phe Leu Pro Glu Asp Thr Pro Gln Gly Arg Asn Ser Asp His
80 85 90
Gly Glu Asp Glu Val Ile Ala Val Ser Glu Lys Val Ile Val Lys
95 100 105
Leu Glu Asp Leu Val Lys Trp Val His Ser Asp Phe Ser Lys Trp
110 115 120
Arg Cys Gly Phe His Ala Gly Pro Val Lys Thr Glu Ala Leu Gly
125 130 135
Arg Asn Gly Gln Lys Glu Ala Leu Leu Lys Tyr Arg Gln Ser Thr
140 145 150
Leu Asn Ser Gly Leu Asn Phe Lys Asp Val Leu Lys Glu Lys Ala
155 160 165
Asp Leu Gly Glu Asp Glu Glu Glu Thr Asn Val Ile Val Leu Ser
170 175 180
Tyr Pro Gln Tyr Cys Arg Tyr Arg Ser Met Leu Lys Arg Ile Gln
185 190 195
Asp Lys Pro Ser Ser Ile Leu Thr Asp Gln Phe Ala Leu Ala Leu
200 205 210
Gly Gly Ile Ala Val Val Ser Arg Asn Pro Gln Ile Leu Tyr Cys
215 220 225
Arg Asp Thr Phe Asp His Pro Thr Leu Ile Glu Asn Glu Ser Ile
230 235 240
Cys Asp Glu Phe Ala Pro Asn Leu Lys Gly Arg Pro Arg Lys Lys
245 250 255
Lys Pro Cys Pro Gln Arg Arg Asp Ser Phe Ser Gly Val Lys Asp
260 265 270
Ser Asn Asn Asn Ser Asp Gly Lys Ala Val Ala Lys Val Lys Cys
275 280 285
Glu Ala Arg Ser Ala Leu Thr Lys Pro Lys Asn Asn His Asn Cys
290 295 300
Lys Lys Val Ser Asn Glu Glu Lys Pro Lys Val Ala Ile Gly Glu
305 310 315
Glu Cys Arg Ala Asp Glu Gln Ala Phe Leu Val Ala Leu Tyr Lys
320 325 330
Tyr Met Lys Glu Arg Lys Thr Pro Ile Glu Arg Ile Pro Tyr Leu
335 340 345
Gly Phe Lys Gln Ile Asn Leu Trp Thr Met Phe Gln Ala Ala Gln
350 355 360
Lys Leu Gly Gly Tyr Glu Thr Ile Thr Ala Arg Arg Gln Trp Lys
365 370 375
His Ile Tyr Asp Glu Leu Gly Gly Asn Pro Gly Ser Thr Ser Ala
380 385 390
Ala Thr Cys Thr Arg Arg His Tyr Glu Arg Leu Ile Leu Pro Tyr
395 400 405
Glu Arg Phe Ile Lys Gly Glu Glu Asp Lys Pro Leu Pro Pro Ile

```
                410                 415                 420
Lys Pro Arg Lys Gln Glu Asn Ser Ser Gln Glu Asn Glu Asn Lys
                425                 430                 435
Thr Lys Val Ser Gly Thr Lys Arg Ile Lys His Glu Ile Pro Lys
                440                 445                 450
Ser Lys Lys Glu Lys Glu Asn Ala Pro Lys Pro Gln Asp Ala Ala
                455                 460                 465
Glu Val Ser Ser Glu Gln Glu Lys Glu Gln Glu Thr Leu Ile Ser
                470                 475                 480
Gln Lys Ser Ile Pro Glu Pro Leu Pro Ala Ala Asp Met Lys Lys
                485                 490                 495
Lys Ile Glu Gly Tyr Gln Glu Phe Ser Ala Lys Pro Leu Ala Ser
                500                 505                 510
Arg Val Asp Pro Glu Lys Asp Asn Glu Thr Asp Gln Gly Ser His
                515                 520                 525
Ser Glu Lys Val Ala Glu Glu Ala Gly Glu Lys Gly Pro Thr Pro
                530                 535                 540
Pro Leu Pro Ser Ala Pro Leu Ala Pro Glu Lys Asp Ser Ala Leu
                545                 550                 555
Val Pro Gly Ala Ser Lys Gln Pro Leu Thr Ser Pro Ser Ala Leu
                560                 565                 570
Val Asp Ser Lys Gln Glu Ser Lys Leu Cys Cys Phe Thr Glu Ser
                575                 580                 585
Pro Glu Ser Glu Pro Gln Glu Ala Ser Phe Pro Thr Thr Gln Pro
                590                 595                 600
Pro Leu Ala Asn Gln Asn Glu Thr Glu Asp Asp Lys Leu Pro Ala
                605                 610                 615
Met Ala Asp Tyr Ile Ala Asn Cys Thr Val Lys Val Asp Gln Leu
                620                 625                 630
Gly Ser Asp Asp Ile His Asn Ala Leu Lys Gln Thr Pro Lys Val
                635                 640                 645
Leu Val Val Gln Ser Phe Asp Met Phe Lys Asp Lys Asp Leu Thr
                650                 655                 660
Gly Pro Met Asn Glu Asn His Gly Leu Asn Tyr Thr Pro Leu Leu
                665                 670                 675
Tyr Ser Arg Gly Asn Pro Gly Ile Met Ser Pro Leu Ala Lys Lys
                680                 685                 690
Lys Leu Leu Ser Gln Val Ser Gly Ala Ser Leu Ser Ser Ser Tyr
                695                 700                 705
Pro Tyr Gly Ser Pro Pro Pro Leu Ile Ser Lys Lys Lys Leu Ile
                710                 715                 720
Ala Arg Asp Asp Leu Cys Ser Ser Leu Ser Gln Thr His His Gly
                725                 730                 735
Gln Ser Thr Asp His Met Ala Val Ser Arg Pro Ser Val Ile Gln
                740                 745                 750
His Val Gln Ser Phe Arg Ser Lys Pro Ser Glu Glu Arg Lys Thr
                755                 760                 765
Ile Asn Asp Ile Phe Lys His Glu Lys Leu Ser Arg Ser Asp Pro
                770                 775                 780
His Arg Cys Ser Phe Ser Lys His His Leu Asn Pro Leu Ala Asp
                785                 790                 795
Ser Tyr Val Leu Lys Gln Glu Ile Gln Glu Gly Lys Asp Lys Leu
                800                 805                 810
Leu Glu Lys Arg Ala Leu Pro His Ser His Met Pro Ser Phe Leu
                815                 820                 825
Ala Asp Phe Tyr Ser Ser Pro His Leu His Ser Leu Tyr Arg His
```

```
                    830                     835                     840
Thr Glu His His Leu His Asn Glu Gln Thr Ser Lys Tyr Pro Ser
                    845                     850                     855
Arg Asp Met Tyr Arg Glu Ser Glu Asn Ser Ser Phe Pro Ser His
                    860                     865                     870
Arg His Gln Glu Lys Leu His Val Asn Tyr Leu Thr Ser Leu His
                    875                     880                     885
Leu Gln Asp Lys Lys Ser Ala Ala Ala Glu Ala Pro Thr Asp Asp
                    890                     895                     900
Gln Pro Thr Asp Leu Ser Leu Pro Lys Asn Pro His Lys Pro Thr
                    905                     910                     915
Gly Lys Val Leu Gly Leu Ala His Ser Thr Thr Gly Pro Gln Glu
                    920                     925                     930
Ser Lys Gly Ile Ser Gln Phe Gln Val Leu Gly Ser Gln Ser Arg
                    935                     940                     945
Asp Cys His Pro Lys Ala Cys Arg Val Ser Pro Met Thr Met Ser
                    950                     955                     960
Gly Pro Lys Lys Tyr Pro Glu Ser Leu Ser Arg Ser Gly Lys Pro
                    965                     970                     975
His His Val Arg Leu Glu Asn Phe Arg Lys Met Glu Gly Met Val
                    980                     985                     990
His Pro Ile Leu His Arg Lys Met Ser Pro Gln Asn Ile Gly Ala
                    995                     1000                    1005
Ala Arg Pro Ile Lys Arg Ser Leu Glu Asp Leu Asp Leu Val Ile
                    1010                    1015                    1020
Ala Gly Lys Lys Ala Arg Ala Val Ser Pro Leu Asp Pro Ser Lys
                    1025                    1030                    1035
Glu Val Ser Gly Lys Glu Lys Ala Ser Glu Gln Glu Ser Glu Gly
                    1040                    1045                    1050
Ser Lys Ala Ala His Gly Gly His Ser Gly Gly Gly Ser Glu Gly
                    1055                    1060                    1065
His Lys Leu Pro Leu Ser Ser Pro Ile Phe Pro Gly Leu Tyr Ser
                    1070                    1075                    1080
Gly Ser Leu Cys Asn Ser Gly Leu Asn Ser Arg Leu Pro Ala Gly
                    1085                    1090                    1095
Tyr Ser His Ser Leu Gln Tyr Leu Lys Asn Gln Thr Val Leu Ser
                    1100                    1105                    1110
Pro Leu Met Gln Pro Leu Ala Phe His Ser Leu Val Met Gln Arg
                    1115                    1120                    1125
Gly Ile Phe Thr Ser Pro Thr Asn Ser Gln Gln Leu Tyr Arg His
                    1130                    1135                    1140
Leu Ala Ala Ala Thr Pro Val Gly Ser Ser Tyr Gly Asp Leu Leu
                    1145                    1150                    1155
His Asn Ser Ile Tyr Pro Leu Ala Ala Ile Asn Pro Gln Ala Ala
                    1160                    1165                    1170
Phe Pro Ser Ser Gln Leu Ser Ser Val His Pro Ser Thr Lys Leu
                    1175                    1180                    1185
```

<210> 10
<211> 1042
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature

<223> Incyte ID No: 878534CD1

<400> 10
Met Ala Ala Met Ala Pro Ala Leu Thr Asp Ala Ala Ala Glu Ala
1   5   10   15
His His Ile Arg Phe Lys Leu Ala Pro Pro Ser Ser Thr Leu Ser
                20   25   30
Pro Gly Ser Ala Glu Asn Asn Gly Asn Ala Asn Ile Leu Ile Ala
                35   40   45
Ala Asn Gly Thr Lys Arg Lys Ala Ile Ala Ala Glu Asp Pro Ser
                50   55   60
Leu Asp Phe Arg Asn Asn Pro Thr Lys Glu Asp Leu Gly Lys Leu
                65   70   75
Gln Pro Leu Val Ala Ser Tyr Leu Cys Ser Asp Val Thr Ser Val
                80   85   90
Pro Ser Lys Glu Ser Leu Lys Leu Gln Gly Val Phe Ser Lys Gln
                95   100   105
Thr Val Leu Lys Ser His Pro Leu Leu Ser Gln Ser Tyr Glu Leu
                110   115   120
Arg Ala Glu Leu Leu Gly Arg Gln Pro Val Leu Glu Phe Ser Leu
                125   130   135
Glu Asn Leu Arg Thr Met Asn Thr Ser Gly Gln Thr Ala Leu Pro
                140   145   150
Gln Ala Pro Val Asn Gly Leu Ala Lys Lys Leu Thr Lys Ser Ser
                155   160   165
Thr His Ser Asp His Asp Asn Ser Thr Ser Leu Asn Gly Gly Lys
                170   175   180
Arg Ala Leu Thr Ser Ser Ala Leu His Gly Gly Glu Met Gly Gly
                185   190   195
Ser Glu Ser Gly Asp Leu Lys Gly Gly Met Thr Asn Cys Thr Leu
                200   205   210
Pro His Arg Ser Leu Asp Val Glu His Thr Ile Leu Tyr Ser Asn
                215   220   225
Asn Ser Thr Ala Asn Lys Ser Ser Val Asn Ser Met Glu Gln Pro
                230   235   240
Ala Leu Gln Gly Ser Ser Arg Leu Ser Pro Gly Thr Asp Ser Ser
                245   250   255
Ser Asn Leu Gly Gly Val Lys Leu Glu Gly Lys Lys Ser Pro Leu
                260   265   270
Ser Ser Ile Leu Phe Ser Ala Leu Asp Ser Asp Thr Arg Ile Thr
                275   280   285
Ala Leu Leu Arg Arg Gln Ala Asp Ile Glu Ser Arg Ala Arg Arg
                290   295   300
Leu Gln Lys Arg Leu Gln Val Val Gln Ala Lys Gln Val Glu Arg
                305   310   315
His Ile Gln His Gln Leu Gly Gly Phe Leu Glu Lys Thr Leu Ser
                320   325   330
Lys Leu Pro Asn Leu Glu Ser Leu Arg Pro Arg Ser Gln Leu Met
                335   340   345
Leu Thr Arg Lys Ala Glu Ala Ala Leu Arg Lys Ala Ala Ser Glu
                350   355   360
Thr Thr Thr Ser Glu Gly Leu Ser Asn Phe Leu Lys Ser Asn Ser
                365   370   375
Ile Ser Glu Glu Leu Glu Arg Phe Thr Ala Ser Gly Ile Ala Asn
                380   385   390
Leu Arg Cys Ser Glu Gln Ala Phe Asp Ser Asp Val Thr Asp Ser

```
                395                 400                 405
Ser Ser Gly Gly Glu Ser Asp Ile Glu Glu Glu Glu Leu Thr Arg
                410                 415                 420
Ala Asp Pro Glu Gln Arg His Val Pro Leu Arg Arg Arg Ser Glu
                425                 430                 435
Trp Lys Trp Ala Ala Asp Arg Ala Ala Ile Val Ser Arg Trp Asn
                440                 445                 450
Trp Leu Gln Ala His Val Ser Asp Leu Glu Tyr Arg Ile Arg Gln
                455                 460                 465
Gln Thr Asp Ile Tyr Lys Gln Ile Arg Ala Asn Lys Gly Leu Ile
                470                 475                 480
Val Leu Gly Glu Val Pro Pro Pro Glu His Thr Thr Asp Leu Phe
                485                 490                 495
Leu Pro Leu Ser Ser Glu Val Lys Thr Asp His Gly Thr Asp Lys
                500                 505                 510
Leu Ile Glu Ser Val Ser Gln Pro Leu Glu Asn His Gly Ala Pro
                515                 520                 525
Ile Ile Gly His Ile Ser Glu Ser Leu Ser Thr Lys Ser Cys Gly
                530                 535                 540
Ala Leu Arg Pro Val Asn Gly Val Ile Asn Thr Leu Gln Pro Val
                545                 550                 555
Leu Ala Asp His Ile Pro Gly Asp Ser Ser Asp Ala Glu Glu Gln
                560                 565                 570
Leu His Lys Lys Gln Arg Leu Asn Leu Val Ser Ser Ser Ser Asp
                575                 580                 585
Gly Thr Cys Val Ala Ala Arg Thr Arg Pro Val Leu Ser Cys Lys
                590                 595                 600
Lys Arg Arg Leu Val Arg Pro Asn Ser Ile Val Pro Leu Ser Lys
                605                 610                 615
Lys Val His Arg Asn Ser Thr Ile Arg Pro Gly Cys Asp Val Asn
                620                 625                 630
Pro Ser Cys Ala Leu Cys Gly Ser Gly Ser Ile Asn Thr Met Pro
                635                 640                 645
Pro Glu Ile His Tyr Glu Ala Pro Leu Leu Glu Arg Leu Ser Gln
                650                 655                 660
Leu Asp Ser Cys Val His Pro Val Leu Ala Phe Pro Asp Asp Val
                665                 670                 675
Pro Thr Ser Leu His Phe Gln Ser Met Leu Lys Ser Gln Trp Gln
                680                 685                 690
Asn Lys Pro Phe Asp Lys Ile Lys Pro Pro Lys Lys Leu Ser Leu
                695                 700                 705
Lys His Arg Ala Pro Met Pro Gly Ser Leu Pro Asp Ser Ala Arg
                710                 715                 720
Lys Asp Arg His Lys Leu Val Ser Ser Phe Leu Thr Thr Ala Met
                725                 730                 735
Leu Lys His His Thr Asp Met Ser Ser Ser Ser Tyr Leu Ala Ala
                740                 745                 750
Thr His His Pro Pro His Ser Pro Leu Val Arg Gln Leu Ser Thr
                755                 760                 765
Ser Ser Asp Ser Pro Ala Pro Ala Ser Ser Ser Ser Gln Val Thr
                770                 775                 780
Ala Ser Thr Ser Gln Gln Pro Val Arg Arg Arg Arg Gly Glu Ser
                785                 790                 795
Ser Phe Asp Ile Asn Asn Ile Val Ile Pro Met Ser Val Ala Ala
                800                 805                 810
Thr Thr Arg Val Glu Lys Leu Gln Tyr Lys Glu Ile Leu Thr Pro
```

```
                815                 820                 825
Ser Trp Arg Glu Val Asp Leu Gln Ser Leu Lys Gly Ser Pro Asp
                830                 835                 840
Glu Glu Asn Glu Glu Ile Glu Asp Leu Ser Asp Ala Ala Phe Ala
                845                 850                 855
Ala Leu His Ala Lys Cys Glu Glu Met Glu Arg Ala Arg Trp Leu
                860                 865                 870
Trp Thr Thr Ser Val Pro Pro Gln Arg Arg Gly Ser Arg Ser Tyr
                875                 880                 885
Arg Ser Ser Asp Gly Arg Thr Thr Pro Gln Leu Gly Ser Ala Asn
                890                 895                 900
Pro Ser Thr Pro Gln Pro Ala Ser Pro Asp Val Ser Ser Ser His
                905                 910                 915
Ser Leu Ser Glu Tyr Ser His Gly Gln Ser Pro Arg Ser Pro Ile
                920                 925                 930
Ser Pro Glu Leu His Ser Ala Pro Leu Thr Pro Val Ala Arg Asp
                935                 940                 945
Thr Leu Arg His Leu Ala Ser Glu Asp Thr Arg Cys Ser Thr Pro
                950                 955                 960
Glu Leu Gly Leu Asp Glu Gln Ser Val Gln Pro Trp Glu Arg Arg
                965                 970                 975
Thr Phe Pro Leu Ala His Ser Pro Gln Ala Glu Cys Glu Asp Gln
                980                 985                 990
Leu Asp Ala Gln Glu Arg Ala Ala Arg Cys Thr Arg Arg Thr Ser
                995                1000                1005
Gly Ser Lys Thr Gly Arg Glu Thr Glu Ala Ala Pro Thr Ser Pro
               1010                1015                1020
Pro Ile Val Pro Leu Lys Ser Arg His Leu Val Ala Ala Ala Thr
               1025                1030                1035
Ala Gln Arg Pro Thr His Arg
               1040
```

<210> 11
<211> 86
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2806157CD1

<400> 11

```
Met Pro Lys Cys Gly Gly Val Arg Val Trp Ile Lys Asp Trp Asn
  1               5                  10                  15
Val Ala Ser Leu Cys Pro Trp Trp Lys Gly Pro Gln Thr Val Val
                 20                  25                  30
Leu Ile Thr Pro Thr Ala Val Asn Val Glu Arg Ile Leu Ala Trp
                 35                  40                  45
Ile His His Asn Arg Val Lys Pro Ala Ala Pro Glu Ser Trp Glu
                 50                  55                  60
Ala Arg Pro Ser Leu Asp Asn Pro Cys Arg Val Thr Leu Lys Lys
                 65                  70                  75
Met Thr Ser Pro Ala Pro Val Thr Pro Arg Ser
                 80                  85
```

<210> 12

<211> 138
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5883626CD1

<400> 12
Met Lys Met Met Val Val Leu Leu Met Leu Ser Ser Leu Ser Arg
1 5 10 15
Leu Leu Gly Leu Met Arg Pro Ser Ser Leu Arg Gln Tyr Leu Asp
20 25 30
Ser Val Pro Leu Pro Pro Cys Gln Glu Gln Gln Pro Lys Ala Ser
35 40 45
Ala Glu Leu Asp His Lys Ala Cys Tyr Leu Cys His Ser Leu Leu
50 55 60
Met Leu Ala Gly Val Val Val Ser Cys Gln Asp Ile Thr Pro Asp
65 70 75
Gln Trp Gly Glu Leu Gln Leu Leu Cys Met Gln Leu Asp Arg His
80 85 90
Ile Ser Thr Gln Ile Arg Glu Ser Pro Gln Ala Met His Arg Thr
95 100 105
Met Leu Lys Asp Leu Ala Thr Gln Thr Tyr Ile Arg Trp Gln Glu
110 115 120
Leu Leu Thr His Cys Gln Pro Gln Ala Gln Tyr Phe Ser Pro Trp
125 130 135
Lys Asp Ile

<210> 13
<211> 805
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2674016CD1

<400> 13
Met Trp Asp Gln Gly Gly Gln Pro Trp Gln Gln Trp Pro Leu Asn
1 5 10 15
Gln Gln Gln Trp Met Gln Ser Phe Gln His Gln Gln Asp Pro Ser
20 25 30
Gln Ile Asp Trp Ala Ala Leu Ala Gln Ala Trp Ile Ala Gln Arg
35 40 45
Glu Ala Ser Gly Gln Gln Ser Met Val Glu Gln Pro Pro Gly Met
50 55 60
Met Pro Asn Gly Gln Asp Met Ser Thr Met Glu Ser Gly Pro Asn
65 70 75
Asn His Gly Asn Phe Gln Gly Asp Ser Asn Phe Asn Arg Met Trp
80 85 90
Gln Pro Glu Trp Gly Met His Gln Gln Pro Pro His Pro Pro Pro
95 100 105
Asp Gln Pro Trp Met Pro Pro Thr Pro Gly Pro Met Asp Ile Val
110 115 120

Pro Pro Ser Glu Asp Ser Asn Ser Gln Asp Ser Gly Glu Phe Ala
125 130 135
Pro Asp Asn Arg His Ile Phe Asn Gln Asn Asn His Asn Phe Gly
140 145 150
Gly Pro Pro Asp Asn Phe Ala Val Gly Pro Val Asn Gln Phe Asp
155 160 165
Tyr Gln His Gly Ala Ala Phe Gly Pro Pro Gln Gly Gly Phe His
170 175 180
Pro Pro Tyr Trp Gln Pro Gly Pro Pro Gly Pro Pro Ala Pro Pro
185 190 195
Gln Asn Arg Arg Glu Arg Pro Ser Ser Phe Arg Asp Arg Gln Arg
200 205 210
Ser Pro Ile Ala Leu Pro Val Lys Gln Glu Pro Pro Gln Ile Asp
215 220 225
Ala Val Lys Arg Arg Thr Leu Pro Ala Trp Ile Arg Glu Gly Leu
230 235 240
Glu Lys Met Glu Arg Glu Lys Gln Lys Lys Leu Glu Lys Glu Arg
245 250 255
Met Glu Gln Gln Arg Ser Gln Leu Ser Lys Lys Glu Lys Lys Ala
260 265 270
Thr Glu Asp Ala Glu Gly Gly Asp Gly Pro Arg Leu Pro Gln Arg
275 280 285
Ser Lys Phe Asp Ser Asp Glu Glu Glu Glu Asp Thr Glu Asn Val
290 295 300
Glu Ala Ala Ser Ser Gly Lys Val Thr Arg Ser Pro Ser Pro Val
305 310 315
Pro Gln Glu Glu His Ser Asp Pro Glu Met Thr Glu Glu Glu Lys
320 325 330
Glu Tyr Gln Met Met Leu Leu Thr Lys Met Leu Leu Thr Glu Ile
335 340 345
Leu Leu Asp Val Thr Asp Glu Glu Ile Tyr Tyr Val Ala Lys Asp
350 355 360
Ala His Arg Lys Ala Thr Lys Ala Pro Ala Lys Gln Leu Ala Gln
365 370 375
Ser Ser Ala Leu Ala Ser Leu Thr Gly Leu Gly Gly Leu Gly Gly
380 385 390
Tyr Gly Ser Gly Asp Ser Glu Asp Glu Arg Ser Asp Arg Gly Ser
395 400 405
Glu Ser Ser Asp Thr Asp Asp Glu Glu Leu Arg His Arg Ile Arg
410 415 420
Gln Lys Gln Glu Ala Phe Trp Arg Lys Glu Lys Glu Gln Gln Leu
425 430 435
Leu His Asp Lys Gln Met Glu Glu Glu Lys Gln Gln Thr Glu Arg
440 445 450
Val Thr Lys Glu Met Asn Glu Phe Ile His Lys Glu Gln Asn Ser
455 460 465
Leu Ser Leu Leu Glu Ala Arg Glu Ala Asp Gly Asp Val Val Asn
470 475 480
Glu Lys Lys Arg Thr Pro Asn Glu Thr Thr Ser Val Leu Glu Pro
485 490 495
Lys Lys Glu His Lys Glu Lys Glu Lys Gln Gly Arg Ser Arg Ser
500 505 510
Gly Ser Ser Ser Ser Gly Ser Ser Ser Ser Asn Ser Arg Thr Ser
515 520 525
Ser Thr Ser Ser Thr Val Ser Ser Ser Ser Tyr Ser Ser Ser Ser
530 535 540

Gly Ser Ser Arg Thr Ser Ser Arg Ser Ser Ser Pro Lys Arg Lys
545 550 555
Lys Arg His Ser Arg Ser Arg Ser Pro Thr Ile Lys Ala Arg Arg
560 565 570
Ser Arg Ser Arg Ser Tyr Ser Arg Arg Ile Lys Ile Glu Ser Asn
575 580 585
Arg Ala Arg Val Lys Ile Arg Asp Arg Arg Arg Ser Asn Arg Asn
590 595 600
Ser Ile Glu Arg Glu Arg Arg Arg Asn Arg Ser Pro Ser Arg Glu
605 610 615
Arg Arg Arg Ser Arg Ser Arg Ser Arg Asp Arg Arg Thr Asn Arg
620 625 630
Ala Ser Arg Ser Arg Ser Arg Asp Arg Arg Lys Ile Asp Asp Gln
635 640 645
Arg Gly Asn Leu Ser Gly Asn Ser His Lys His Lys Gly Glu Ala
650 655 660
Lys Glu Gln Glu Arg Lys Lys Glu Arg Ser Arg Ser Ile Asp Lys
665 670 675
Asp Arg Lys Lys Lys Asp Lys Glu Arg Glu Arg Glu Gln Asp Lys
680 685 690
Arg Lys Glu Lys Gln Lys Arg Glu Glu Lys Asp Phe Lys Phe Ser
695 700 705
Ser Gln Asp Asp Arg Leu Lys Arg Lys Arg Glu Ser Glu Arg Thr
710 715 720
Phe Ser Arg Ser Gly Ser Ile Ser Val Lys Ile Ile Arg His Asp
725 730 735
Ser Arg Gln Asp Ser Lys Lys Ser Thr Thr Lys Asp Ser Lys Lys
740 745 750
His Ser Gly Ser Asp Ser Ser Gly Arg Ser Ser Ser Glu Ser Pro
755 760 765
Gly Ser Ser Lys Glu Lys Lys Ala Lys Lys Pro Lys His Ser Arg
770 775 780
Ser Arg Ser Val Glu Lys Ser Pro Arg Ser Gly Lys Lys Ala Ser
785 790 795
Arg Lys His Lys Ser Lys Ser Arg Ser Arg
800 805

<210> 14
<211> 426
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5994159CD1

<400> 14
Met Val Gly Ala Ala His Arg Ala Gln Ala Val Phe Thr Val Val
1 5 10 15
Ser Ser Glu Leu Lys Gly Met Cys Phe His Leu Pro Met Arg Thr
20 25 30
Ala Pro Ser Val Ser Val Trp Leu Glu Thr Cys Pro Ala Ser Leu
35 40 45
Leu Ser Val Leu Leu Ala Pro Val Arg Pro Pro His Arg Arg Ile
50 55 60
Ala Val Leu Val Phe Gln Ala Asp Gly Ser Val Ser Cys Lys Arg

```
                65                  70                  75
Thr Asp Cys Val Asp Ser Cys Pro His Pro Ile Arg Ile Pro Gly
                80                  85                  90
Gln Cys Cys Pro Asp Cys Ser Ala Gly Cys Thr Tyr Thr Gly Arg
                95                  100                 105
Ile Phe Tyr Asn Asn Glu Thr Phe Pro Ser Val Leu Asp Pro Cys
                110                 115                 120
Leu Ser Cys Ile Cys Leu Leu Gly Ser Val Ala Cys Ser Pro Val
                125                 130                 135
Asp Cys Pro Ile Thr Cys Thr Tyr Pro Phe His Pro Asp Gly Glu
                140                 145                 150
Cys Cys Pro Val Cys Arg Asp Cys Asn Tyr Glu Gly Arg Lys Val
                155                 160                 165
Ala Asn Gly Gln Val Phe Thr Leu Asp Asp Glu Pro Cys Thr Arg
                170                 175                 180
Cys Thr Cys Gln Leu Gly Glu Val Ser Cys Glu Lys Val Pro Cys
                185                 190                 195
Gln Arg Ala Cys Ala Asp Pro Ala Leu Leu Pro Gly Asp Cys Cys
                200                 205                 210
Ser Ser Cys Pro Asp Ser Leu Ser Pro Leu Glu Glu Lys Gln Gly
                215                 220                 225
Leu Ser Pro His Gly Asn Val Ala Phe Ser Lys Ala Gly Arg Ser
                230                 235                 240
Leu His Gly Asp Thr Glu Ala Pro Val Asn Cys Ser Ser Cys Pro
                245                 250                 255
Gly Pro Pro Thr Ala Ser Pro Ser Arg Pro Val Leu His Leu Leu
                260                 265                 270
Gln Leu Leu Leu Arg Thr Asn Leu Met Lys Thr Gln Thr Leu Pro
                275                 280                 285
Thr Ser Pro Ala Gly Ala His Gly Pro His Ser Leu Ala Leu Gly
                290                 295                 300
Leu Thr Ala Thr Phe Pro Gly Glu Pro Gly Ala Ser Pro Arg Leu
                305                 310                 315
Ser Pro Gly Pro Ser Thr Pro Pro Gly Ala Pro Thr Leu Pro Leu
                320                 325                 330
Ala Ser Pro Gly Ala Pro Gln Pro Pro Pro Val Thr Pro Glu Arg
                335                 340                 345
Ser Phe Ser Ala Ser Gly Ala Gln Ile Val Ser Arg Trp Pro Pro
                350                 355                 360
Leu Pro Gly Thr Leu Leu Thr Glu Ala Ser Ala Leu Ser Met Met
                365                 370                 375
Asp Pro Ser Pro Ser Lys Thr Pro Ile Thr Leu Leu Gly Pro Arg
                380                 385                 390
Val Leu Ser Pro Thr Thr Ser Arg Leu Ser Thr Ala Leu Ala Ala
                395                 400                 405
Thr Thr His Pro Gly Pro Gln Gln Pro Pro Val Gly Ala Ser Arg
                410                 415                 420
Gly Glu Glu Ser Thr Met
                425
```

<210> 15
<211> 267
<212> PRT
<213> Homo sapiens

<220>

<221> misc_feature
<223> Incyte ID No: 2457335CD1

<400> 15
Met Tyr Leu Arg Arg Ala Val Ser Lys Thr Leu Ala Leu Pro Leu
1   5   10   15
Arg Ala Pro Pro Asn Pro Ala Pro Leu Gly Lys Asp Ala Ser Leu
                20   25   30
Arg Arg Met Ser Ser Asn Arg Phe Pro Gly Ser Ser Gly Ser Asn
                35   40   45
Met Ile Tyr Tyr Leu Val Val Gly Val Thr Val Ser Ala Gly Gly
                50   55   60
Tyr Tyr Ala Tyr Lys Thr Val Thr Ser Asp Gln Ala Lys His Thr
                65   70   75
Glu His Lys Thr Asn Leu Lys Glu Lys Thr Lys Ala Glu Ile His
                80   85   90
Pro Phe Gln Gly Glu Lys Glu Asn Val Ala Glu Thr Glu Lys Ala
                95   100   105
Ser Ser Glu Ala Pro Glu Glu Leu Ile Val Glu Ala Glu Val Val
                110   115   120
Asp Ala Glu Glu Ser Pro Ser Ala Thr Val Val Val Ile Lys Glu
                125   130   135
Ala Ser Ala Cys Pro Gly His Val Glu Ala Ala Pro Glu Thr Thr
                140   145   150
Ala Val Ser Ala Glu Thr Gly Pro Glu Val Thr Asp Ala Ala Ala
                155   160   165
Arg Glu Thr Thr Glu Val Asn Pro Glu Thr Thr Pro Glu Val Thr
                170   175   180
Asn Ala Ala Leu Asp Glu Ala Val Thr Ile Asp Asn Asp Lys Asp
                185   190   195
Thr Thr Lys Asn Glu Thr Ser Asp Glu Tyr Ala Glu Leu Glu Glu
                200   205   210
Glu Asn Ser Pro Ala Glu Ser Glu Ser Ser Ala Gly Asp Asp Leu
                215   220   225
Gln Glu Glu Ala Ser Val Gly Ser Glu Ala Ala Ser Ala Gln Gly
                230   235   240
Asn Leu Gln Pro Val Asp Ile Ser Ala Thr Asn Ala Ile Gly Cys
                245   250   255
Leu Ile Ser Ala Leu Val Phe Leu Val His Leu Val
                260   265

<210> 16
<211> 928
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2267802CD1

<400> 16
Met Glu Gly Ala Gly Glu Asn Ala Pro Glu Ser Ser Ser Ser Ala
1   5   10   15
Pro Gly Ser Glu Glu Ser Ala Arg Asp Pro Gln Val Pro Pro Pro
                20   25   30
Glu Glu Glu Ser Gly Asp Cys Ala Arg Ser Leu Glu Ala Val Pro

```
                35                       40                      45
Lys Lys Leu Cys Gly Tyr Leu Ser Lys Phe Gly Gly Lys Gly Pro
                50                       55                      60
Ile Arg Gly Trp Lys Ser Arg Trp Phe Phe Tyr Asp Glu Arg Lys
                65                       70                      75
Cys Gln Leu Tyr Tyr Ser Arg Thr Ala Gln Asp Ala Asn Pro Leu
                80                       85                      90
Asp Ser Ile Asp Leu Ser Ser Ala Val Phe Asp Cys Lys Ala Asp
                95                      100                     105
Ala Glu Glu Gly Ile Phe Glu Ile Lys Thr Pro Ser Arg Val Ile
               110                      115                     120
Thr Leu Lys Ala Ala Thr Lys Gln Ala Met Leu Tyr Trp Leu Gln
               125                      130                     135
Gln Leu Gln Met Lys Arg Trp Glu Phe His Asn Ser Pro Pro Ala
               140                      145                     150
Pro Pro Ala Thr Pro Asp Ala Ala Leu Ala Gly Asn Gly Pro Val
               155                      160                     165
Leu His Leu Glu Leu Gly Gln Glu Glu Ala Glu Leu Glu Glu Phe
               170                      175                     180
Leu Cys Pro Val Lys Thr Pro Pro Gly Leu Val Gly Val Ala Ala
               185                      190                     195
Ala Leu Gln Pro Phe Pro Ala Leu Gln Asn Ile Ser Leu Lys His
               200                      205                     210
Leu Gly Thr Glu Ile Gln Asn Thr Met His Asn Ile Arg Gly Asn
               215                      220                     225
Lys Gln Ala Gln Gly Thr Gly His Glu Pro Pro Gly Glu Asp Ser
               230                      235                     240
Thr Gln Ser Gly Glu Pro Gln Arg Glu Glu Gln Pro Ser Ala Ser
               245                      250                     255
Asp Ala Ser Thr Pro Val Arg Glu Pro Glu Asp Ser Pro Lys Pro
               260                      265                     270
Ala Pro Lys Pro Ser Leu Thr Ile Ser Phe Ala Gln Lys Ala Lys
               275                      280                     285
Arg Gln Asn Asn Thr Phe Pro Phe Phe Ser Glu Gly Ile Thr Arg
               290                      295                     300
Asn Arg Thr Ala Gln Glu Lys Val Ala Ala Leu Glu Gln Gln Val
               305                      310                     315
Leu Met Leu Thr Lys Glu Leu Lys Ser Gln Lys Glu Leu Val Lys
               320                      325                     330
Ile Leu His Lys Ala Leu Glu Ala Ala Gln Gln Glu Lys Arg Ala
               335                      340                     345
Ser Ser Ala Tyr Leu Ala Ala Ala Glu Asp Lys Asp Arg Leu Glu
               350                      355                     360
Leu Val Arg His Lys Val Arg Gln Ile Ala Glu Leu Gly Arg Arg
               365                      370                     375
Val Glu Ala Leu Glu Gln Glu Arg Glu Ser Leu Ala His Thr Ala
               380                      385                     390
Ser Leu Arg Glu Gln Gln Val Gln Glu Leu Gln Gln His Val Gln
               395                      400                     405
Leu Leu Met Asp Lys Asn His Ala Glu Gln Gln Val Ile Cys Lys
               410                      415                     420
Leu Ser Glu Lys Val Thr Gln Asp Phe Thr His Pro Pro Asp Gln
               425                      430                     435
Ser Pro Leu Arg Pro Asp Ala Ala Asn Arg Asp Phe Leu Ser Gln
               440                      445                     450
Gln Gly Lys Ile Glu His Leu Lys Asp Asp Met Glu Ala Tyr Arg
```

```
                455                 460                 465
Thr Gln Asn Cys Phe Leu Asn Ser Glu Ile His Gln Val Thr Lys
                470                 475                 480
Ile Trp Arg Lys Val Ala Glu Lys Glu Lys Ala Leu Leu Thr Lys
                485                 490                 495
Cys Ala Tyr Leu Gln Ala Arg Asn Cys Gln Val Glu Ser Lys Tyr
                500                 505                 510
Leu Ala Gly Leu Arg Arg Leu Gln Glu Ala Leu Gly Asp Glu Ala
                515                 520                 525
Ser Glu Cys Ser Glu Leu Leu Arg Gln Leu Val Gln Glu Ala Leu
                530                 535                 540
Gln Trp Glu Ala Gly Glu Ala Ser Ser Asp Ser Ile Glu Leu Ser
                545                 550                 555
Pro Ile Ser Lys Tyr Asp Glu Tyr Gly Phe Leu Thr Val Pro Asp
                560                 565                 570
Tyr Glu Val Glu Asp Leu Lys Leu Leu Ala Lys Ile Gln Ala Leu
                575                 580                 585
Glu Ser Arg Ser His His Leu Leu Gly Leu Glu Ala Val Asp Arg
                590                 595                 600
Pro Leu Arg Glu Arg Trp Ala Ala Leu Gly Asp Leu Val Pro Ser
                605                 610                 615
Ala Glu Leu Lys Gln Leu Leu Arg Ala Gly Val Pro Arg Glu His
                620                 625                 630
Arg Pro Arg Val Trp Arg Trp Leu Val His Leu Arg Val Gln His
                635                 640                 645
Leu His Thr Pro Gly Cys Tyr Gln Glu Leu Leu Ser Arg Gly Gln
                650                 655                 660
Ala Arg Glu His Pro Ala Ala Arg Gln Ile Glu Leu Asp Leu Asn
                665                 670                 675
Arg Thr Phe Pro Asn Asn Lys His Phe Thr Cys Pro Thr Ser Ser
                680                 685                 690
Phe Pro Asp Lys Leu Arg Arg Val Leu Leu Ala Phe Ser Trp Gln
                695                 700                 705
Asn Pro Thr Ile Gly Tyr Cys Gln Gly Leu Asn Arg Leu Ala Ala
                710                 715                 720
Ile Ala Leu Leu Val Leu Glu Glu Glu Glu Ser Ala Phe Trp Cys
                725                 730                 735
Leu Val Ala Ile Val Glu Thr Ile Met Pro Ala Asp Tyr Tyr Cys
                740                 745                 750
Asn Thr Leu Thr Ala Ser Gln Val Asp Gln Arg Val Leu Gln Asp
                755                 760                 765
Leu Leu Ser Glu Lys Leu Pro Arg Leu Met Ala His Leu Gly Gln
                770                 775                 780
His His Val Asp Leu Ser Leu Val Thr Phe Asn Trp Phe Leu Val
                785                 790                 795
Val Phe Ala Asp Ser Leu Ile Ser Asn Ile Leu Leu Arg Val Trp
                800                 805                 810
Asp Ala Phe Leu Tyr Glu Gly Thr Lys Val Val Phe Arg Tyr Ala
                815                 820                 825
Leu Ala Ile Phe Lys Tyr Asn Glu Lys Glu Ile Leu Arg Leu Gln
                830                 835                 840
Asn Gly Leu Glu Ile Tyr Gln Tyr Leu Arg Phe Phe Thr Lys Thr
                845                 850                 855
Ile Ser Asn Ser Arg Lys Leu Met Asn Ile Ala Phe Asn Asp Met
                860                 865                 870
Asn Pro Phe Arg Met Lys Gln Leu Arg Gln Leu Arg Met Val His
```

```
                875                 880                 885
Arg Glu Arg Leu Glu Ala Glu Leu Arg Glu Leu Glu Gln Leu Lys
                890                 895                 900
Ala Glu Tyr Leu Glu Arg Arg Ala Ser Arg Arg Arg Ala Val Ser
                905                 910                 915
Glu Gly Cys Ala Ser Glu Asp Glu Val Glu Gly Glu Ala
                920                 925

<210> 17
<211> 684
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3212060CD1

<400> 17
Met Trp Val Leu Leu Arg Ser Gly Tyr Pro Leu Arg Ile Leu Leu
  1               5                  10                  15
Pro Leu Arg Gly Glu Trp Met Gly Arg Arg Gly Leu Pro Arg Asn
                 20                  25                  30
Leu Ala Pro Gly Pro Pro Arg Arg Arg Tyr Arg Lys Glu Thr Leu
                 35                  40                  45
Gln Ala Leu Asp Met Pro Val Leu Pro Val Thr Ala Thr Glu Ile
                 50                  55                  60
Arg Gln Tyr Leu Arg Gly His Gly Ile Pro Phe Gln Asp Gly His
                 65                  70                  75
Ser Cys Leu Arg Ala Leu Ser Pro Phe Ala Glu Ser Ser Gln Leu
                 80                  85                  90
Lys Gly Gln Thr Gly Val Thr Thr Ser Phe Ser Leu Phe Ile Asp
                 95                 100                 105
Lys Thr Thr Gly His Phe Leu Cys Met Thr Ser Leu Ala Glu Gly
                110                 115                 120
Ser Trp Glu Asp Phe Gln Ala Ser Val Glu Gly Arg Gly Asp Gly
                125                 130                 135
Ala Arg Glu Gly Phe Leu Leu Ser Lys Ala Pro Glu Phe Glu Asp
                140                 145                 150
Ser Glu Glu Val Arg Arg Ile Trp Asn Arg Ala Ile Pro Leu Trp
                155                 160                 165
Glu Leu Pro Asp Gln Glu Glu Val Gln Leu Ala Asp Thr Met Phe
                170                 175                 180
Gly Leu Thr Lys Val Thr Asp Asp Thr Leu Lys Arg Phe Ser Val
                185                 190                 195
Arg Tyr Leu Arg Pro Ala Arg Ser Leu Val Phe Pro Trp Phe Ser
                200                 205                 210
Pro Gly Gly Ser Gly Leu Arg Gly Leu Lys Leu Leu Glu Ala Lys
                215                 220                 225
Cys Gln Gly Asp Gly Val Ser Tyr Glu Glu Thr Thr Ile Pro Arg
                230                 235                 240
Pro Ser Ala Tyr His Asn Leu Phe Gly Leu Pro Leu Ile Ser Arg
                245                 250                 255
Arg Asp Ala Glu Val Val Leu Thr Ser Arg Glu Leu Asp Ser Leu
                260                 265                 270
Ala Leu Asn Gln Ser Thr Gly Leu Pro Thr Leu Thr Leu Pro Arg
                275                 280                 285
```

```
Gly Thr Thr Cys Leu Pro Pro Ala Leu Leu Pro Tyr Leu Glu Gln
                290                 295                 300
Phe Arg Arg Ile Val Phe Trp Leu Gly Asp Asp Leu Arg Ser Trp
                305                 310                 315
Glu Ala Ala Lys Leu Phe Ala Arg Lys Leu Asn Pro Lys Arg Cys
                320                 325                 330
Phe Leu Val Arg Pro Gly Asp Gln Gln Pro Arg Pro Leu Glu Ala
                335                 340                 345
Leu Asn Gly Gly Phe Asn Leu Ser Arg Ile Leu Arg Thr Ala Leu
                350                 355                 360
Pro Ala Trp His Lys Ser Ile Val Ser Phe Arg Gln Leu Arg Glu
                365                 370                 375
Glu Val Leu Gly Glu Leu Ser Asn Val Glu Gln Ala Ala Gly Leu
                380                 385                 390
Arg Trp Ser Arg Phe Pro Asp Leu Asn Arg Ile Leu Lys Gly His
                395                 400                 405
Arg Lys Gly Glu Leu Thr Val Phe Thr Gly Pro Thr Gly Ser Gly
                410                 415                 420
Lys Thr Thr Phe Ile Ser Glu Tyr Ala Leu Asp Leu Cys Ser Gln
                425                 430                 435
Gly Val Asn Thr Leu Trp Gly Ser Phe Glu Ile Ser Asn Val Arg
                440                 445                 450
Leu Ala Arg Val Met Leu Thr Gln Phe Ala Glu Gly Arg Leu Glu
                455                 460                 465
Asp Gln Leu Asp Lys Tyr Asp His Trp Ala Asp Arg Phe Glu Asp
                470                 475                 480
Leu Pro Leu Tyr Phe Met Thr Phe His Gly Gln Gln Ser Ile Arg
                485                 490                 495
Thr Val Ile Asp Thr Met Gln His Ala Val Tyr Val Tyr Asp Ile
                500                 505                 510
Cys His Val Ile Ile Asp Asn Leu Gln Phe Met Met Gly His Glu
                515                 520                 525
Gln Leu Ser Thr Asp Arg Ile Ala Ala Gln Asp Tyr Ile Ile Gly
                530                 535                 540
Val Phe Arg Lys Phe Ala Thr Asp Asn Asn Cys His Val Thr Leu
                545                 550                 555
Val Ile His Pro Arg Lys Glu Asp Asp Asp Lys Glu Leu Gln Thr
                560                 565                 570
Ala Ser Ile Phe Gly Ser Ala Lys Ala Ser Gln Glu Ala Asp Asn
                575                 580                 585
Val Leu Ile Leu Gln Asp Arg Lys Leu Val Thr Gly Pro Gly Lys
                590                 595                 600
Arg Tyr Leu Gln Val Ser Lys Asn Arg Phe Asp Gly Asp Val Gly
                605                 610                 615
Val Phe Pro Leu Glu Phe Asn Lys Asn Ser Leu Thr Phe Ser Ile
                620                 625                 630
Pro Pro Lys Asn Lys Ala Arg Leu Lys Lys Ile Lys Asp Asp Thr
                635                 640                 645
Gly Pro Val Ala Lys Lys Pro Ser Ser Gly Lys Lys Gly Ala Thr
                650                 655                 660
Thr Gln Asn Ser Glu Ile Cys Ser Gly Gln Ala Pro Thr Pro Asp
                665                 670                 675
Gln Pro Asp Thr Ser Lys Arg Ser Lys
                680
```

<210> 18

<211> 267
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3121069CD1

<400> 18
Met Thr Lys Thr Ala Leu Leu Lys Leu Phe Val Ala Ile Val Ile
1 5 10 15
Thr Phe Ile Leu Ile Leu Pro Glu Tyr Phe Lys Thr Pro Lys Glu
20 25 30
Arg Thr Leu Glu Leu Ser Cys Leu Glu Val Cys Leu Gln Ser Asn
35 40 45
Phe Thr Tyr Ser Leu Ser Ser Leu Asn Phe Ser Phe Val Thr Phe
50 55 60
Leu Gln Pro Val Arg Glu Thr Gln Ile Ile Met Arg Ile Phe Leu
65 70 75
Asn Pro Ser Asn Phe Arg Asn Phe Thr Arg Thr Cys Gln Asp Ile
80 85 90
Thr Val Leu Ile Arg Arg Gly Ser Met Glu Val Lys Ala Asn Asp
95 100 105
Phe His Ser Pro Cys Gln His Phe Asn Phe Ser Val Ala Pro Leu
110 115 120
Val Asp His Leu Glu Glu Tyr Asn Thr Thr Cys His Leu Lys Asn
125 130 135
His Thr Gly Arg Ser Thr Ile Met Glu Asp Glu Pro Ser Lys Glu
140 145 150
Lys Ser Ile Asn Tyr Thr Cys Arg Ile Met Glu Tyr Pro Asn Asp
155 160 165
Cys Ile His Ile Ser Leu His Leu Glu Met Asp Ile Lys Asn Ile
170 175 180
Thr Cys Ser Met Lys Ile Thr Trp Tyr Ile Leu Val Leu Leu Val
185 190 195
Phe Ile Phe Leu Ile Ile Leu Thr Ile Arg Lys Ile Leu Glu Gly
200 205 210
Gln Arg Arg Val Gln Lys Trp Gln Ser His Arg Asp Lys Pro Thr
215 220 225
Ser Val Leu Leu Arg Gly Ser Asp Ser Glu Lys Leu Arg Ala Leu
230 235 240
Asn Val Gln Val Leu Ser Glu Thr Thr Gln Arg Leu Pro Leu Asp
245 250 255
Gln Val Gln Glu Val Leu Pro Pro Ile Pro Glu Leu
260 265

<210> 19
<211> 537
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3280626CD1

<400> 19

Met Ala Asp Asn Leu Asp Glu Phe Ile Glu Glu Gln Lys Ala Arg
1 5 10 15
Leu Ala Glu Asp Lys Ala Glu Leu Glu Ser Asp Pro Pro Tyr Met
20 25 30
Glu Met Lys Gly Lys Leu Ser Ala Lys Leu Ser Glu Asn Ser Lys
35 40 45
Ile Leu Ile Ser Met Ala Lys Glu Asn Ile Pro Pro Asn Ser Gln
50 55 60
Gln Thr Arg Gly Ser Leu Gly Ile Asp Tyr Gly Leu Ser Leu Pro
65 70 75
Leu Gly Glu Asp Tyr Glu Arg Lys Lys His Lys Leu Lys Glu Glu
80 85 90
Leu Arg Gln Asp Tyr Arg Arg Tyr Leu Thr Gln Glu Arg Leu Lys
95 100 105
Leu Glu Arg Asn Lys Glu Tyr Asn Gln Phe Leu Arg Gly Lys Glu
110 115 120
Glu Ser Ser Glu Lys Phe Arg Gln Val Glu Lys Ser Thr Glu Pro
125 130 135
Lys Ser Gln Arg Asn Lys Lys Pro Ile Gly Gln Val Lys Pro Asp
140 145 150
Leu Thr Ser Gln Ile Gln Thr Ser Cys Glu Asn Ser Glu Gly Pro
155 160 165
Arg Lys Asp Val Leu Thr Pro Ser Glu Ala Tyr Glu Glu Leu Leu
170 175 180
Asn Gln Arg Arg Leu Glu Glu Asp Arg Tyr Arg Gln Leu Asp Asp
185 190 195
Glu Ile Glu Leu Arg Asn Arg Arg Ile Ile Lys Lys Ala Asn Glu
200 205 210
Glu Val Gly Ile Ser Asn Leu Lys His Gln Arg Phe Ala Ser Lys
215 220 225
Ala Gly Ile Pro Asp Arg Arg Phe His Arg Phe Asn Glu Asp Arg
230 235 240
Val Phe Asp Arg Arg Tyr His Arg Pro Asp Gln Asp Pro Glu Val
245 250 255
Ser Glu Glu Met Asp Glu Arg Phe Arg Tyr Glu Ser Asp Phe Asp
260 265 270
Arg Arg Leu Ser Arg Val Tyr Thr Asn Asp Arg Met His Arg Asn
275 280 285
Lys Arg Gly Asn Met Pro Pro Met Glu His Asp Gly Asp Val Ile
290 295 300
Glu Gln Ser Asn Ile Arg Ile Ser Ser Ala Glu Asn Lys Ser Ala
305 310 315
Pro Asp Asn Glu Thr Ser Lys Ser Ala Asn Gln Asp Thr Cys Ser
320 325 330
Pro Phe Ala Gly Met Leu Phe Gly Gly Glu Asp Arg Glu Leu Ile
335 340 345
Gln Arg Arg Lys Glu Lys Tyr Arg Leu Glu Leu Leu Glu Gln Met
350 355 360
Ala Glu Gln Gln Arg Asn Lys Arg Arg Glu Lys Asp Leu Glu Leu
365 370 375
Arg Val Ala Ala Ser Gly Ala Gln Asp Pro Glu Lys Ser Pro Asp
380 385 390
Arg Leu Lys Gln Phe Ser Val Ala Pro Arg His Phe Glu Glu Met
395 400 405
Ile Pro Pro Glu Arg Pro Arg Ile Ala Phe Gln Thr Pro Leu Pro
410 415 420

Pro Leu Ser Ala Pro Ser Val Pro Pro Ile Pro Ser Val His Pro
425 430 435
Val Pro Ser Gln Asn Glu Asp Leu Arg Ser Gly Leu Ser Ser Ala
440 445 450
Leu Gly Glu Met Val Ser Pro Arg Ile Ala Pro Leu Pro Pro Pro
455 460 465
Pro Leu Leu Pro Pro Leu Ala Thr Asn Tyr Arg Thr Pro Tyr Asp
470 475 480
Asp Ala Tyr Tyr Phe Tyr Gly Ser Arg Asn Thr Phe Asp Pro Ser
485 490 495
Leu Ala Tyr Tyr Gly Ser Gly Met Met Gly Val Gln Pro Ala Ala
500 505 510
Tyr Val Ser Ala Pro Val Thr His Gln Leu Ala Gln Pro Val Val
515 520 525
Val Ser Pro Cys His Pro Gly Trp Ser Thr Met Leu
530 535

<210> 20
<211> 312
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 484404CD1

<400> 20
Met Trp Ser Glu Gly Arg Tyr Glu Tyr Glu Arg Ile Pro Arg Glu
1 5 10 15
Arg Ala Pro Pro Arg Ser His Pro Ser Asp Glu Ser Gly Tyr Arg
20 25 30
Trp Thr Arg Asp Asp His Ser Ala Ser Arg Gln Pro Glu Tyr Arg
35 40 45
Asp Met Arg Asp Gly Phe Arg Arg Lys Ser Phe Tyr Ser Ser His
50 55 60
Tyr Ala Arg Glu Arg Ser Pro Tyr Lys Arg Asp Asn Thr Phe Phe
65 70 75
Arg Glu Ser Pro Val Gly Arg Lys Asp Ser Pro His Ser Arg Ser
80 85 90
Gly Ser Ser Val Ser Ser Arg Ser Tyr Ser Pro Glu Arg Ser Lys
95 100 105
Ser Tyr Ser Phe His Gln Ser Gln His Arg Lys Ser Val Arg Pro
110 115 120
Gly Ala Ser Tyr Lys Arg Gln Asn Glu Gly Asn Pro Glu Arg Asp
125 130 135
Lys Glu Arg Pro Val Gln Ser Leu Lys Thr Ser Arg Asp Thr Ser
140 145 150
Pro Ser Ser Gly Ser Ala Val Ser Ser Ser Lys Val Leu Asp Lys
155 160 165
Pro Ser Arg Leu Thr Glu Lys Glu Leu Ala Glu Ala Ala Ser Lys
170 175 180
Trp Ala Ala Glu Lys Leu Glu Lys Ser Asp Glu Ser Asn Leu Pro
185 190 195
Glu Ile Ser Glu Tyr Glu Ala Gly Ser Thr Ala Pro Leu Phe Thr
200 205 210
Asp Gln Pro Glu Glu Pro Glu Ser Asn Thr Thr His Gly Ile Glu

```
                215                 220                 225
Leu Phe Glu Asp Ser Gln Leu Thr Thr Arg Ser Lys Ala Ile Ala
                230                 235                 240
Ser Lys Thr Lys Glu Ile Glu Gln Val Tyr Arg Gln Asp Cys Glu
                245                 250                 255
Thr Phe Gly Met Val Val Lys Met Leu Ile Glu Lys Asp Pro Ser
                260                 265                 270
Leu Glu Lys Ser Ile Gln Phe Ala Leu Arg Gln Asn Leu His Glu
                275                 280                 285
Ile Gly Glu Arg Cys Val Glu Glu Leu Lys His Phe Ile Ala Glu
                290                 295                 300
Tyr Asp Thr Ser Thr Gln Asp Phe Gly Glu Pro Phe
                305                 310

<210> 21
<211> 1400
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2830063CD1

<400> 21
Met Met Ala Ser Phe Gln Arg Ser Asn Ser His Asp Lys Val Arg
  1               5                  10                  15
Arg Ile Val Ala Glu Glu Gly Arg Thr Ala Arg Asn Leu Ile Ala
                 20                  25                  30
Trp Ser Val Pro Leu Glu Ser Lys Asp Asp Asp Gly Lys Pro Lys
                 35                  40                  45
Cys Gln Thr Gly Gly Lys Ser Lys Arg Thr Ile Gln Gly Thr His
                 50                  55                  60
Lys Thr Thr Lys Gln Ser Thr Ala Val Asp Cys Lys Ile Thr Ser
                 65                  70                  75
Ser Thr Thr Gly Asp Lys His Phe Asp Lys Ser Pro Thr Lys Thr
                 80                  85                  90
Arg His Pro Arg Lys Ile Asp Leu Arg Ala Arg Tyr Trp Ala Phe
                 95                 100                 105
Leu Phe Asp Asn Leu Arg Arg Ala Val Asp Glu Ile Tyr Val Thr
                110                 115                 120
Cys Glu Ser Asp Gln Ser Val Val Glu Cys Lys Glu Val Leu Met
                125                 130                 135
Met Leu Asp Asn Tyr Val Arg Asp Phe Lys Ala Leu Ile Asp Trp
                140                 145                 150
Ile Gln Leu Gln Glu Lys Leu Glu Lys Thr Asp Ala Gln Ser Arg
                155                 160                 165
Pro Thr Ser Leu Ala Trp Glu Val Lys Lys Met Ser Pro Gly Arg
                170                 175                 180
His Val Ile Pro Ser Pro Ser Thr Asp Arg Ile Asn Val Thr Ser
                185                 190                 195
Asn Ala Arg Arg Ser Leu Asn Phe Gly Gly Ser Thr Gly Thr Val
                200                 205                 210
Pro Ala Pro Arg Leu Ala Pro Thr Gly Val Ser Trp Ala Asp Lys
                215                 220                 225
Val Lys Ala His His Thr Gly Ser Thr Ala Ser Ser Glu Ile Thr
                230                 235                 240
```

```
Pro Ala Gln Ser Cys Pro Pro Met Thr Val Gln Lys Ala Ser Arg
                245                 250                 255
Lys Asn Glu Arg Lys Asp Ala Glu Gly Trp Glu Thr Val Gln Arg
                260                 265                 270
Gly Arg Pro Ile Arg Ser Arg Ser Thr Ala Val Met Pro Lys Val
                275                 280                 285
Ser Leu Ala Thr Glu Ala Thr Arg Ser Lys Asp Asp Ser Asp Lys
                290                 295                 300
Glu Asn Val Cys Leu Leu Pro Asp Glu Ser Ile Gln Lys Gly Gln
                305                 310                 315
Phe Val Gly Asp Gly Thr Ser Asn Thr Ile Glu Ser His Pro Lys
                320                 325                 330
Asp Ser Leu His Ser Cys Asp His Pro Leu Ala Glu Lys Thr Gln
                335                 340                 345
Phe Thr Val Ser Thr Leu Asp Asp Val Lys Asn Ser Gly Ser Ile
                350                 355                 360
Arg Asp Asn Tyr Val Arg Thr Ser Glu Ile Ser Ala Val His Ile
                365                 370                 375
Asp Thr Glu Cys Val Ser Val Met Leu Gln Ala Gly Thr Pro Pro
                380                 385                 390
Leu Gln Val Asn Glu Glu Lys Phe Pro Ala Glu Lys Ala Arg Ile
                395                 400                 405
Glu Asn Glu Met Asp Pro Ser Asp Ile Ser Asn Ser Met Ala Glu
                410                 415                 420
Val Leu Ala Lys Lys Glu Glu Leu Ala Asp Arg Leu Glu Lys Ala
                425                 430                 435
Asn Glu Glu Ala Ile Ala Ser Ala Ile Ala Glu Glu Glu Gln Leu
                440                 445                 450
Thr Arg Glu Ile Glu Ala Glu Glu Asn Asn Asp Ile Asn Ile Glu
                455                 460                 465
Thr Asp Asn Asp Ser Asp Phe Ser Ala Ser Met Gly Ser Gly Ser
                470                 475                 480
Val Ser Phe Cys Gly Met Ser Met Asp Trp Asn Asp Val Leu Ala
                485                 490                 495
Asp Tyr Glu Ala Arg Glu Ser Trp Arg Gln Asn Thr Ser Trp Gly
                500                 505                 510
Asp Ile Val Glu Glu Glu Pro Ala Arg Pro Pro Gly His Gly Ile
                515                 520                 525
His Met His Glu Lys Leu Ser Ser Pro Ser Arg Lys Arg Thr Ile
                530                 535                 540
Ala Glu Ser Lys Lys Lys His Glu Glu Lys Gln Met Lys Ala Gln
                545                 550                 555
Gln Leu Arg Glu Lys Leu Arg Glu Glu Lys Thr Leu Lys Leu Gln
                560                 565                 570
Lys Leu Leu Glu Arg Glu Lys Asp Val Arg Lys Trp Lys Glu Glu
                575                 580                 585
Leu Leu Asp Gln Arg Arg Arg Met Met Glu Glu Lys Leu Leu His
                590                 595                 600
Ala Glu Phe Lys Arg Glu Val Gln Leu Gln Ala Ile Val Lys Lys
                605                 610                 615
Ala Gln Glu Glu Glu Ala Lys Val Asn Glu Ile Ala Phe Ile Asn
                620                 625                 630
Thr Leu Glu Ala Gln Asn Lys Arg His Asp Val Leu Ser Lys Leu
                635                 640                 645
Lys Glu Tyr Glu Gln Arg Leu Asn Glu Leu Gln Glu Glu Arg Gln
                650                 655                 660
```

Arg Arg Gln Glu Glu Lys Gln Ala Arg Asp Glu Ala Val Gln Glu
665 670 675
Arg Lys Arg Ala Leu Glu Ala Glu Arg Gln Ala Arg Val Glu Glu
680 685 690
Leu Leu Met Lys Arg Lys Glu Gln Glu Ala Arg Ile Glu Gln Gln
695 700 705
Arg Gln Glu Lys Glu Lys Ala Arg Glu Asp Ala Ala Arg Glu Arg
710 715 720
Ala Arg Asp Arg Glu Glu Arg Leu Ala Ala Leu Thr Ala Ala Gln
725 730 735
Gln Glu Ala Met Glu Glu Leu Gln Lys Lys Ile Gln Leu Lys His
740 745 750
Asp Glu Ser Ile Arg Arg His Met Glu Gln Ile Glu Gln Arg Lys
755 760 765
Glu Lys Ala Ala Glu Leu Ser Ser Gly Arg His Ala Asn Thr Asp
770 775 780
Tyr Ala Pro Lys Leu Thr Pro Tyr Glu Arg Lys Lys Gln Cys Ser
785 790 795
Leu Cys Asn Val Leu Ile Ser Ser Glu Val Tyr Leu Phe Ser His
800 805 810
Val Lys Gly Arg Lys His Gln Gln Ala Val Arg Glu Asn Thr Ser
815 820 825
Ile Gln Gly Arg Glu Leu Ser Asp Glu Glu Val Glu His Leu Ser
830 835 840
Leu Lys Lys Tyr Ile Ile Asp Ile Val Val Glu Ser Thr Ala Pro
845 850 855
Ala Glu Ala Leu Lys Asp Gly Glu Glu Arg Gln Lys Asn Lys Lys
860 865 870
Lys Ala Lys Lys Ile Lys Ala Arg Met Asn Phe Arg Ala Lys Glu
875 880 885
Tyr Glu Ser Leu Met Glu Thr Lys Asn Ser Gly Ser Asp Ser Pro
890 895 900
Tyr Lys Ala Lys Leu Gln Arg Leu Ala Lys Asp Leu Leu Lys Gln
905 910 915
Val Gln Val Gln Asp Ser Gly Ser Trp Ala Asn Asn Lys Val Ser
920 925 930
Ala Leu Asp Arg Thr Leu Gly Glu Ile Thr Arg Ile Leu Glu Lys
935 940 945
Glu Asn Val Ala Asp Gln Ile Ala Phe Gln Ala Ala Gly Gly Leu
950 955 960
Thr Ala Leu Glu His Ile Leu Gln Ala Val Val Pro Ala Thr Asn
965 970 975
Val Asn Thr Val Leu Arg Ile Pro Pro Lys Ser Leu Cys Asn Ala
980 985 990
Ile Asn Val Tyr Asn Leu Thr Cys Asn Asn Cys Ser Glu Asn Cys
995 1000 1005
Ser Asp Val Leu Phe Ser Asn Lys Ile Thr Phe Leu Met Asp Leu
1010 1015 1020
Leu Ile His Gln Leu Thr Val Tyr Val Pro Asp Glu Asn Asn Thr
1025 1030 1035
Ile Leu Gly Arg Asn Thr Asn Lys Gln Val Phe Glu Gly Leu Thr
1040 1045 1050
Thr Gly Leu Leu Lys Val Ser Ala Val Val Leu Gly Cys Leu Ile
1055 1060 1065
Ala Asn Arg Pro Asp Gly Asn Cys Gln Pro Ala Thr Pro Lys Ile
1070 1075 1080

Pro Thr Gln Glu Met Lys Asn Lys Thr Ser Gln Gly Asp Pro Phe
1085 1090 1095
Asn Asn Arg Val Gln Asp Leu Ile Ser Tyr Val Val Asn Met Gly
1100 1105 1110
Leu Ile Asp Lys Leu Cys Ala Cys Phe Leu Ser Val Gln Gly Pro
1115 1120 1125
Val Asp Glu Asn Pro Lys Met Ala Ile Phe Leu Gln His Ala Ala
1130 1135 1140
Gly Leu Leu His Ala Met Cys Thr Leu Cys Phe Ala Val Thr Gly
1145 1150 1155
Arg Ser Tyr Ser Ile Phe Asp Asn Asn Arg Gln Asp Pro Thr Gly
1160 1165 1170
Leu Thr Ala Ala Leu Gln Ala Thr Asp Leu Ala Gly Val Leu His
1175 1180 1185
Met Leu Tyr Cys Val Leu Phe His Gly Thr Ile Leu Asp Pro Ser
1190 1195 1200
Thr Ala Ser Pro Lys Glu Asn Tyr Thr Gln Asn Thr Ile Gln Val
1205 1210 1215
Ala Ile Gln Ser Leu Arg Phe Phe Asn Ser Phe Ala Ala Leu His
1220 1225 1230
Leu Pro Ala Phe Gln Ser Ile Val Gly Ala Glu Gly Leu Ser Leu
1235 1240 1245
Ala Phe Arg His Met Ala Ser Ser Leu Leu Gly His Cys Ser Gln
1250 1255 1260
Val Ser Cys Glu Ser Leu Leu His Glu Val Ile Val Cys Val Gly
1265 1270 1275
Tyr Phe Thr Val Asn His Pro Asp Asn Gln Val Ile Val Gln Ser
1280 1285 1290
Gly Arg His Pro Thr Val Leu Gln Lys Leu Cys Gln Leu Pro Phe
1295 1300 1305
Gln Tyr Phe Ser Asp Pro Arg Leu Ile Lys Val Leu Phe Pro Ser
1310 1315 1320
Leu Ile Ala Ala Cys Tyr Asn Asn His Gln Asn Lys Ile Ile Leu
1325 1330 1335
Glu Gln Glu Met Ser Cys Val Leu Leu Ala Thr Phe Ile Gln Asp
1340 1345 1350
Leu Ala Gln Thr Pro Gly Gln Ala Glu Asn Gln Pro Tyr Gln Pro
1355 1360 1365
Lys Gly Lys Cys Leu Gly Ser Gln Asp Tyr Leu Glu Leu Ala Asn
1370 1375 1380
Arg Phe Pro Gln Gln Ala Trp Glu Glu Ala Arg Gln Phe Phe Leu
1385 1390 1395
Lys Lys Glu Lys Lys
1400

<210> 22
<211> 1384
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7506096CD1

<400> 22
Met Glu Ser Ser Ser Ser Asp Tyr Tyr Asn Lys Asp Asn Glu Glu

```
1               5                   10                  15
Glu Ser Leu Leu Ala Asn Val Ala Ser Leu Arg His Glu Leu Lys
                20                  25                  30
Ile Thr Glu Trp Ser Leu Gln Ser Leu Gly Glu Glu Leu Ser Ser
                35                  40                  45
Val Ser Pro Ser Glu Asn Ser Asp Tyr Ala Pro Asn Pro Ser Arg
                50                  55                  60
Ser Glu Lys Leu Ile Leu Asp Val Gln Pro Ser His Pro Gly Leu
                65                  70                  75
Leu Asn Tyr Ser Pro Tyr Glu Asn Val Cys Lys Ile Ser Gly Ser
                80                  85                  90
Ser Thr Asp Phe Gln Lys Lys Pro Arg Asp Lys Met Phe Ser Ser
                95                  100                 105
Ser Ala Pro Val Asp Gln Glu Ile Lys Ser Leu Arg Glu Lys Leu
                110                 115                 120
Asn Lys Leu Arg Gln Gln Asn Ala Cys Leu Val Thr Gln Asn His
                125                 130                 135
Ser Leu Met Thr Lys Phe Glu Ser Ile His Phe Glu Leu Thr Gln
                140                 145                 150
Ser Arg Ala Lys Val Ser Met Leu Glu Ser Ala Gln Gln Gln Ala
                155                 160                 165
Ala Ser Val Pro Ile Leu Glu Glu Gln Ile Ile Asn Leu Glu Ala
                170                 175                 180
Glu Val Ser Ala Gln Asp Lys Val Leu Arg Glu Ala Glu Asn Lys
                185                 190                 195
Leu Glu Gln Ser Gln Lys Met Val Ile Glu Lys Glu Gln Ser Leu
                200                 205                 210
Gln Glu Ser Lys Glu Glu Cys Ile Lys Leu Lys Val Asp Leu Leu
                215                 220                 225
Glu Gln Thr Lys Gln Gly Lys Arg Ala Glu Arg Gln Arg Asn Glu
                230                 235                 240
Ala Leu Tyr Asn Ala Glu Glu Leu Ser Lys Ala Phe Gln Gln Tyr
                245                 250                 255
Lys Lys Lys Val Ala Glu Lys Leu Glu Lys Val Gln Ala Glu Glu
                260                 265                 270
Glu Ile Leu Glu Arg Asn Leu Thr Asn Cys Glu Lys Glu Asn Lys
                275                 280                 285
Arg Leu Gln Glu Arg Cys Gly Leu Tyr Lys Ser Glu Leu Glu Ile
                290                 295                 300
Leu Lys Glu Lys Leu Arg Gln Leu Lys Glu Glu Asn Asn Asn Gly
                305                 310                 315
Lys Glu Lys Leu Arg Ile Met Ala Val Lys Asn Ser Glu Val Met
                320                 325                 330
Ala Gln Leu Thr Glu Ser Arg Gln Ser Ile Leu Lys Leu Glu Ser
                335                 340                 345
Glu Leu Glu Asn Lys Asp Glu Ile Leu Arg Asp Lys Phe Ser Leu
                350                 355                 360
Met Asn Glu Asn Arg Glu Leu Lys Val Arg Val Ala Ala Gln Asn
                365                 370                 375
Glu Arg Leu Asp Leu Cys Gln Gln Glu Ile Glu Ser Ser Arg Val
                380                 385                 390
Glu Leu Arg Ser Leu Glu Lys Ile Ile Ser Gln Leu Pro Leu Lys
                395                 400                 405
Arg Glu Leu Phe Gly Phe Lys Ser Tyr Leu Ser Lys Tyr Gln Met
                410                 415                 420
Ser Ser Phe Ser Asn Lys Glu Asp Arg Cys Ile Gly Cys Cys Glu
```

```
                    425                       430                       435
Ala Asn Lys Leu Val Ile Ser Glu Leu Arg Ile Lys Leu Ala Ile
                    440                       445                       450
Lys Glu Ala Glu Ile Gln Lys Leu His Ala Asn Leu Thr Ala Asn
                    455                       460                       465
Gln Leu Ser Gln Ser Leu Ile Thr Cys Asn Asp Ser Gln Glu Ser
                    470                       475                       480
Ser Lys Leu Ser Ser Leu Glu Thr Glu Pro Val Lys Leu Gly Gly
                    485                       490                       495
His Gln Val Ala Glu Ser Val Lys Asp Gln Asn Gln His Thr Met
                    500                       505                       510
Asn Lys Gln Tyr Glu Lys Glu Arg Gln Arg Leu Val Thr Gly Ile
                    515                       520                       525
Glu Glu Leu Arg Thr Lys Leu Ile Gln Ile Glu Ala Glu Asn Ser
                    530                       535                       540
Asp Leu Lys Val Asn Met Ala His Arg Thr Ser Gln Phe Gln Leu
                    545                       550                       555
Ile Gln Glu Glu Leu Leu Glu Lys Ala Ser Asn Ser Ser Lys Leu
                    560                       565                       570
Glu Ser Glu Met Thr Lys Lys Cys Ser Gln Leu Leu Thr Leu Glu
                    575                       580                       585
Lys Gln Leu Glu Glu Lys Ile Val Ala Tyr Ser Ser Ile Ala Ala
                    590                       595                       600
Lys Asn Ala Glu Leu Glu Gln Glu Leu Met Glu Lys Asn Glu Lys
                    605                       610                       615
Ile Arg Ser Leu Glu Thr Asn Ile Asn Thr Glu His Glu Lys Ile
                    620                       625                       630
Cys Leu Ala Phe Glu Lys Ala Lys Lys Ile His Leu Glu Gln His
                    635                       640                       645
Lys Glu Met Glu Lys Gln Ile Glu Arg Val Arg Gln Leu Asp Ser
                    650                       655                       660
Ala Leu Glu Ile Cys Lys Glu Glu Leu Val Leu His Leu Asn Gln
                    665                       670                       675
Leu Glu Gly Asn Lys Glu Lys Phe Glu Lys Gln Leu Lys Lys Lys
                    680                       685                       690
Ser Glu Glu Val Tyr Cys Leu Gln Lys Glu Leu Lys Ile Lys Asn
                    695                       700                       705
His Ser Leu Gln Glu Thr Ser Glu Gln Asn Val Ile Leu Gln His
                    710                       715                       720
Thr Leu Gln Gln Gln Gln Gln Met Leu Gln Gln Glu Thr Ile Arg
                    725                       730                       735
Asn Gly Glu Leu Glu Asp Thr Gln Thr Lys Leu Glu Lys Gln Val
                    740                       745                       750
Ser Lys Leu Glu Gln Glu Leu Gln Lys Gln Arg Glu Ser Ser Ala
                    755                       760                       765
Glu Lys Leu Arg Lys Met Glu Glu Lys Cys Glu Ser Ala Ala His
                    770                       775                       780
Glu Ala Asp Leu Lys Arg Gln Lys Val Ile Glu Leu Thr Gly Thr
                    785                       790                       795
Ala Arg Gln Val Lys Ile Glu Met Asp Gln Tyr Lys Glu Glu Leu
                    800                       805                       810
Ser Lys Met Glu Lys Glu Ile Met His Leu Lys Arg Asp Gly Glu
                    815                       820                       825
Asn Lys Ala Met His Leu Ser Gln Leu Asp Met Ile Leu Asp Gln
                    830                       835                       840
Thr Lys Thr Glu Leu Glu Lys Lys Thr Asn Ala Val Lys Glu Leu
```

```
                845                 850                 855
Glu Lys Leu Gln His Ser Thr Glu Thr Glu Leu Thr Glu Ala Leu
                860                 865                 870
Gln Lys Arg Glu Val Leu Glu Thr Glu Leu Gln Asn Ala His Gly
                875                 880                 885
Glu Leu Lys Ser Thr Leu Arg Gln Leu Gln Glu Leu Arg Asp Val
                890                 895                 900
Leu Gln Lys Ala Gln Leu Ser Leu Glu Glu Lys Tyr Thr Thr Ile
                905                 910                 915
Lys Asp Leu Thr Ala Glu Leu Arg Glu Cys Lys Met Glu Ile Glu
                920                 925                 930
Asp Lys Lys Gln Glu Leu Leu Glu Met Asp Gln Ala Leu Lys Glu
                935                 940                 945
Arg Asn Trp Glu Leu Lys Gln Arg Ala Ala Gln Val Thr His Leu
                950                 955                 960
Asp Met Thr Ile Arg Glu His Arg Gly Glu Met Glu Gln Lys Ile
                965                 970                 975
Ile Lys Leu Glu Gly Thr Leu Glu Lys Ser Glu Leu Glu Leu Lys
                980                 985                 990
Glu Cys Asn Lys Gln Ile Glu Ser Leu Asn Asp Lys Leu Gln Asn
                995                 1000                1005
Ala Lys Glu Gln Leu Arg Glu Lys Glu Phe Ile Met Leu Gln Asn
                1010                1015                1020
Glu Gln Glu Ile Ser Gln Leu Lys Lys Glu Ile Glu Arg Thr Gln
                1025                1030                1035
Gln Arg Met Lys Glu Met Glu Ser Val Met Lys Glu Gln Glu Gln
                1040                1045                1050
Tyr Ile Ala Thr Gln Tyr Lys Glu Ala Ile Asp Leu Gly Gln Glu
                1055                1060                1065
Leu Arg Leu Thr Arg Glu Gln Val Gln Asn Ser His Thr Glu Leu
                1070                1075                1080
Ala Glu Ala Arg His Gln Gln Val Gln Ala Gln Arg Glu Ile Glu
                1085                1090                1095
Arg Leu Ser Ser Glu Leu Glu Asp Met Lys Gln Leu Ser Lys Glu
                1100                1105                1110
Lys Asp Ala His Gly Asn His Leu Ala Glu Glu Leu Gly Ala Ser
                1115                1120                1125
Lys Val Arg Glu Ala His Leu Glu Ala Arg Met Gln Ala Glu Ile
                1130                1135                1140
Lys Lys Leu Ser Ala Glu Val Glu Ser Leu Lys Glu Ala Tyr His
                1145                1150                1155
Met Glu Met Ile Ser His Gln Glu Asn His Ala Lys Trp Lys Ile
                1160                1165                1170
Ser Ala Asp Ser Gln Lys Ser Ser Val Gln Gln Leu Asn Glu Gln
                1175                1180                1185
Leu Glu Lys Ala Lys Leu Glu Leu Glu Glu Ala Gln Asp Thr Val
                1190                1195                1200
Ser Asn Leu His Gln Gln Val Gln Asp Arg Asn Glu Val Ile Glu
                1205                1210                1215
Ala Ala Asn Glu Ala Leu Leu Thr Lys Glu Ser Glu Leu Thr Arg
                1220                1225                1230
Leu Gln Ala Lys Ile Ser Gly His Glu Lys Ala Glu Asp Ile Lys
                1235                1240                1245
Phe Leu Pro Ala Pro Phe Thr Ser Pro Thr Glu Ile Met Pro Asp
                1250                1255                1260
Val Gln Asp Pro Lys Phe Ala Lys Cys Phe His Thr Ser Phe Ser
```

```
                1265                1270                1275
Lys Cys Thr Lys Leu Arg Arg Ser Ile Ser Ala Ser Asp Leu Thr
                1280                1285                1290
Phe Lys Ile His Gly Asp Glu Asp Leu Ser Glu Glu Leu Leu Gln
                1295                1300                1305
Asp Leu Lys Lys Met Gln Leu Glu Gln Pro Ser Thr Leu Glu Glu
                1310                1315                1320
Ser His Lys Asn Leu Thr Tyr Thr Gln Pro Asp Ser Phe Lys Pro
                1325                1330                1335
Leu Thr Tyr Asn Leu Glu Ala Asp Ser Ser Glu Asn Asn Asp Phe
                1340                1345                1350
Asn Thr Leu Ser Gly Met Leu Arg Tyr Ile Asn Lys Glu Val Arg
                1355                1360                1365
Leu Leu Lys Lys Ser Ser Met Gln Thr Gly Ala Gly Leu Asn Gln
                1370                1375                1380
Gly Glu Asn Val
```

<210> 23
<211> 787
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7505914CD1

<400> 23

```
Met Trp Asp Gln Gly Gly Gln Pro Trp Gln Gln Trp Pro Leu Asn
  1               5                  10                  15
Gln Gln Gln Trp Met Gln Ser Phe Gln His Gln Gln Asp Pro Ser
                 20                  25                  30
Gln Ile Asp Trp Ala Ala Leu Ala Gln Ala Trp Ile Ala Gln Arg
                 35                  40                  45
Glu Ala Ser Gly Gln Gln Ser Met Val Glu Gln Pro Pro Gly Met
                 50                  55                  60
Met Pro Asn Gly Gln Asp Met Ser Thr Met Glu Ser Gly Pro Asn
                 65                  70                  75
Asn His Gly Asn Phe Gln Gly Asp Ser Asn Phe Asn Arg Met Trp
                 80                  85                  90
Gln Pro Glu Trp Gly Met His Gln Gln Pro Pro His Pro Pro Pro
                 95                 100                 105
Asp Gln Pro Trp Met Pro Pro Thr Pro Gly Pro Met Asp Ile Val
                110                 115                 120
Pro Pro Ser Glu Asp Ser Asn Ser Gln Asp Ser Gly Glu Phe Ala
                125                 130                 135
Pro Asp Asn Arg His Ile Phe Asn Gln Asn Asn His Asn Phe Gly
                140                 145                 150
Gly Pro Pro Asp Asn Phe Ala Val Gly Pro Val Asn Gln Phe Asp
                155                 160                 165
Tyr Gln His Gly Ala Ala Phe Gly Pro Pro Gln Gly Gly Phe His
                170                 175                 180
Pro Pro Tyr Trp Gln Pro Gly Pro Pro Gly Pro Pro Ala Pro Pro
                185                 190                 195
Gln Asn Arg Arg Glu Arg Pro Ser Ser Phe Arg Asp Arg Gln Arg
                200                 205                 210
```

Ser Pro Ile Ala Leu Pro Val Lys Gln Glu Pro Pro Gln Ile Asp
215 220 225
Ala Val Lys Arg Arg Thr Leu Pro Ala Trp Ile Arg Glu Gly Leu
230 235 240
Glu Lys Met Glu Arg Glu Lys Gln Lys Lys Leu Glu Lys Glu Arg
245 250 255
Met Glu Gln Gln Arg Ser Gln Leu Ser Lys Lys Glu Lys Lys Ala
260 265 270
Thr Glu Asp Ala Glu Gly Gly Asp Gly Pro Arg Leu Pro Gln Arg
275 280 285
Ser Lys Phe Asp Ser Asp Glu Glu Glu Glu Asp Thr Glu Asn Val
290 295 300
Glu Ala Ala Ser Ser Gly Lys Val Thr Arg Ser Pro Ser Pro Val
305 310 315
Pro Gln Glu Glu His Ser Asp Pro Glu Met Thr Glu Glu Glu Lys
320 325 330
Glu Tyr Gln Met Met Leu Leu Thr Lys Met Leu Leu Thr Glu Ile
335 340 345
Leu Leu Asp Val Thr Asp Glu Glu Ile Tyr Tyr Val Ala Lys Asp
350 355 360
Ala His Arg Lys Ala Thr Lys Gly Gly Leu Gly Gly Tyr Gly Ser
365 370 375
Gly Asp Ser Glu Asp Glu Arg Ser Asp Arg Gly Ser Glu Ser Ser
380 385 390
Asp Thr Asp Asp Glu Glu Leu Arg His Arg Ile Arg Gln Lys Gln
395 400 405
Glu Ala Phe Trp Arg Lys Glu Lys Glu Gln Gln Leu Leu His Asp
410 415 420
Lys Gln Met Glu Glu Glu Lys Gln Gln Thr Glu Arg Val Thr Lys
425 430 435
Glu Met Asn Glu Phe Ile His Lys Glu Gln Asn Ser Leu Ser Leu
440 445 450
Leu Glu Ala Arg Glu Ala Asp Gly Asp Val Val Asn Glu Lys Lys
455 460 465
Arg Thr Pro Asn Glu Thr Thr Ser Val Leu Glu Pro Lys Lys Glu
470 475 480
His Lys Glu Lys Glu Lys Gln Gly Arg Ser Arg Ser Gly Ser Ser
485 490 495
Ser Ser Gly Ser Ser Ser Ser Asn Ser Arg Thr Ser Ser Thr Ser
500 505 510
Ser Thr Val Ser Ser Ser Ser Tyr Ser Ser Ser Ser Gly Ser Ser
515 520 525
Arg Thr Ser Ser Arg Ser Ser Ser Pro Lys Arg Lys Lys Arg His
530 535 540
Ser Arg Ser Arg Ser Pro Thr Ile Lys Ala Arg Arg Ser Arg Ser
545 550 555
Arg Ser Tyr Ser Arg Arg Ile Lys Ile Glu Ser Asn Arg Ala Arg
560 565 570
Val Lys Ile Arg Asp Arg Arg Arg Ser Asn Arg Asn Ser Ile Glu
575 580 585
Arg Glu Arg Arg Arg Asn Arg Ser Pro Ser Arg Glu Arg Arg Arg
590 595 600
Ser Arg Ser Arg Ser Arg Asp Arg Arg Thr Asn Arg Ala Ser Arg
605 610 615
Ser Arg Ser Arg Asp Arg Arg Lys Ile Asp Asp Gln Arg Gly Asn
620 625 630

Leu Ser Gly Asn Ser His Lys His Lys Gly Glu Ala Lys Glu Gln
635 640 645
Glu Arg Lys Lys Glu Arg Ser Arg Ser Ile Asp Lys Asp Arg Lys
650 655 660
Lys Lys Asp Lys Glu Arg Glu Arg Glu Gln Asp Lys Arg Lys Glu
665 670 675
Lys Gln Lys Arg Glu Glu Lys Asp Phe Lys Phe Ser Ser Gln Asp
680 685 690
Asp Arg Leu Lys Arg Lys Arg Glu Ser Glu Arg Thr Phe Ser Arg
695 700 705
Ser Gly Ser Ile Ser Val Lys Ile Ile Arg His Asp Ser Arg Gln
710 715 720
Asp Ser Lys Lys Ser Thr Thr Lys Asp Ser Lys Lys His Ser Gly
725 730 735
Ser Asp Ser Ser Gly Arg Ser Ser Ser Glu Ser Pro Gly Ser Ser
740 745 750
Lys Glu Lys Lys Ala Lys Lys Pro Lys His Ser Arg Ser Arg Ser
755 760 765
Val Glu Lys Ser Gln Arg Ser Gly Lys Lys Ala Ser Arg Lys His
770 775 780
Lys Ser Lys Ser Arg Ser Arg
785

<210> 24
<211> 3332
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 71230017CB1

<400> 24
gccgggcttt gggttctggg cctctgccgc tctctggccc taagtgctga gctgccggga 60
acggcagctt ctgacgctgg gccattggac gctgcggaac caggcttctt cactttgagt 120
ttccgccgcg aagcgccagt ccgggccgag gagggagcct ttactacttc tccctggttt 180
cattcatgtt ctgaggaggg tgtgagaagg aaccatggat cccacagcct tggtggaagc 240
cattgtggaa gaagtggcct gtcccatctg tatgaccttc ctgagggagc ccatgagcat 300
tgactgtggc cacagcttct gccacagctg tctctctgga ctctgggaga tcccaggaga 360
atcccagaac tggggttaca cctgtcccct ctgtcgagct cctgtccagc caaggaacct 420
gcggcctaat tggcagctgg ccaatgttgt agaaaaagtc cgtctgctaa ggctacatcc 480
aggaatgggg ctgaagggtg acctgtgtga gcgccatggg gaaaagctga agatgttctg 540
caaagaggat gtcttgataa tgtgtgaggc ctgcagccag tcccagagc atgaggccca 600
cagtgttgtg ccaatggagg atgttgcctg ggagtacaag tgggaacttc atgaggccct 660
cgaacatctg aagaaagagc aagaagaggc ctggaagctt gaagttggtg aaaggaaacg 720
aactgccacc tggaagatac aggtggaaac ccgaaaacag agtattgtat gggagtttga 780
aaaataccag cgattactag agaaaaagca gccaccacat cggcagctgg gggcagaggt 840
agcagcagct ctggccagcc tacagcggga ggcagcggag accatgcaga aactggagtt 900
gaaccatagc gagctcatcc agcagagcca ggtcctgtgg aggatgattg cagagttgaa 960
agagaggtcg cagaggcctg tccgctggat gttgcaggat attcaggaag tgttaaacag 1020
gagcaaatct tggagcttgc agcagccaga accaatctcc ctggagttga agacagattg 1080
ccgtgtgctg ggcctaagag agatcctgaa gacttatgca gctgatgtgc gcttggatcc 1140
agatactgct tactcccgtc tcatcgtgtc tgaggacaga aaacgtgtgc actatggaga 1200
caccaaccag aaactgccag acaatcctga gagattttac cgctataata tcgtcctggg 1260
aagccagtgc atctcctcag gccggcacta ctgggaggtg gaggtgggag acaggtctga 1320
gtggggcctg ggagtatgta agcaaaatgt agaccggaag gaggtggtct acttatcccc 1380

ccactatgga ttctgggtga taaggctgag gaagggaaat gagtaccgag caggcaccga 1440
tgagtaccca atcctgtcct tgccggtccc tcctcgccgg gtgggaatct tcgtggatta 1500
tgaggcccat gacatttctt tctacaatgt gactgactgt ggctcccaca tcttcacttt 1560
cccccgctat cccttccctg ggcgcctcct gccctatttt agtccttgct acagcattgg 1620
aaccaacaac actgctcctc tggccatctg ctccctggat ggggaggact aagaaagcta 1680
ccaccctaac cacagaggct tggaattggg cctggccccc atggggcttg gaggaccgag 1740
ccactgacag gtatcccctg aaactgagct gagcccagta tccaaggatt cctctgtctg 1800
atcctttggt ctttgctacc aggctgaagt ctgtcatgaa accacttatt ttaaaaagca 1860
gaggcccagt caaatgagca ttgcatccca tgagggaagc acgacagggc tgatggtgag 1920
gatcagagca gttctaaggt gactcgttgg ggtaaggatc aggactttgt ccatgtagta 1980
gccaaccacc ctcttccctg attcccgtcc ggtgtcacag ttcagtcagt gaggatgatg 2040
aagtagatac agtcttcagg acaccattag atgggctttc ccaataggcc aaaaaaatgc 2100
tgcgcatacc cagagctggt tgttgtgctg aggccagtca gaggatgctt cccctgaggt 2160
ttgctataac taagcaacct ttatgtgact ctcaccttct gacctcctgg caagagaaat 2220
tcagtgcagc aggggacac agacctgccc aagccacccc actgccgttc cctctctgag 2280
cacaagctgg gcaaatcact gtcccttgga ctccagtaga ccagtgtcct agtcttgcct 2340
tttttctcta agtggcagga tcagaaaacc tgcgagcttt agtttgtatt ttcactttat 2400
gaatgaggaa actgaaatgg ccttaaggga gcaagttatt tctttttttt tgacacggag 2460
tctcgctctg ttgcccaggc tggagtgcag tggcacgatc tcggctcact gcaggctctg 2520
cctcctgggt tcacgccatt ctcctgcctc agcttcccga gtagctggga ctacaggcgc 2580
ccaccacgac gcctggctca ttttttgta tttttagtag agacggggtt tcaccatgtt 2640
agctaggatg gtctcgatct cctgacctca tgatccgccc tcctcagcct cccacagtgc 2700
tgggattaga ggcatgagcc actgcgcccg gcccctggag caagttattt cttacaaagc 2760
tgctgaaggt aagattatca aaattataaa gcatttttca cactcaagtg aaacaaggtt 2820
gacaaactca cttcgcaggt cacatgccta tacatcactt attatatttg ggtctgaaac 2880
ttctcacatg tttgggaggt tttatgtgtc ctcattggga aaatgggtgt aattcagcat 2940
aaaacctcat atgattgtcc tgcctcatgg agctgttgta tagatcccag atccatccca 3000
tgatttgttc ctgtctgagg catagaggca ggcaagccgt ggattttgca catggtgact 3060
ttcccactgt gccatgatac agtctgcatc ttatagcagt gcctttgtct cagggcctct 3120
gctggcagtc tagacctttt gggcagaaag gagcttcaaa tggctgtgat aaggaatatt 3180
aaaaattgtg tttctacttt aattgtattg gctgttcatg tatgtaggag ttaaaatagg 3240
ccaaactgga gaaataaacg cattctgtcc accatgaaaa aaaaaaaaaa aaaaaaaaaa 3300
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aa 3332

<210> 25
<211> 4410
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3125036CB1

<400> 25
atggaatcta gttcatcaga ctactataat aaagacaatg aagaggaaag tttgcttgca 60
aatgttgctt ccttaagaca tgaactgaag ataacagaat ggagtttgca gagtttaggg 120
gaagagttat ccagtgttag tccaagtgaa aattctgatt atgcccctaa tccttcaagg 180
tctgaaaagc taattttgga tgttcagcct agccaccctg gacttttgaa ttattcacct 240
tatgaaaacg tctgtaaaat atctggtagc agcactgatt ttcaaaaaaa gccaagagat 300
aagatgtttt catcttctgc ccctgtggat caggagatta aaagccttcg agagaaacta 360
aataaactta ggcaacagaa tgcttgtttg gtcacacaga atcattcctt aatgactaaa 420
tttgaatcta ttcactttga attaacacag tcaagagcaa aagtttctat gcttgagtct 480
gctcaacagc aggcagccag tgtcccaatc ttagaagaac agattataaa tttggaagca 540
gaggtttcag ctcaagataa agttttgaga gaggcagaaa ataagctgga acagagccag 600
aaaatggtaa ttgaaaagga acagagtttg caggagtcca aagaggaatg tataaaatta 660
aaggtggact tacttgaaca aaccaaacaa ggaaaaagag ctgaacgaca aaggaatgaa 720

gcactatata atgccgaaga gctgagtaaa gctttccaac aatataaaaa aaaagtggct 780
gaaaaactgg aaaaggtaaa aggcagttgt gcaaattcag tgttttgtat tactgtctat 840
attccaacag taaaggttca agctgaagaa gaaatattag agagaaatct aactaactgt 900
gaaaaagaaa ataaaaggct acaagaaagg tgtggtctat ataaaagtga acttgaaatt 960
ctgaaagaga aattaaggca gttaaaagaa gaaaataaca acggaaaaga aaaattaagg 1020
atcatggcag tgaaaaattc agaagtcatg gcacaactaa ctgaatctag acaaagtatt 1080
ttgaagctag agagtgagtt agagaacaaa gacgaaatac ttagagacaa attttcttta 1140
atgaatgaaa accgagaatt aaaggtccgt gttgcagcac agaatgagcg actagattta 1200
tgtcaacaag aaattgaaag ttcaagggta gaactaagaa gtttggaaaa gattatatcc 1260
cagttgccat taaaaagaga attatttggc tttaaatcat atctttctaa ataccagatg 1320
agtagcttct caaacaagga agaccgttgc attggctgct gtgaggcaaa taaattggtg 1380
atttcggaat tgagaattaa gcttgcaata aaagaggcag aaattcaaaa gcttcatgca 1440
aacctgactg caaatcagtt atctcagagt cttattactt gtaatgacag ccaagaaagt 1500
agcaaattaa gtagtttaga aacagaacct gtaaagctag gtggtcatca agtagcagaa 1560
agcgtaaaag atcaaaatca acatactatg aacaagcaat atgaaaaaga gaggcaaaga 1620
cttgttactg gaatagaaga actacgtact aagctgatac aaatagaagc tgaaaattct 1680
gatttgaagg ttaacatggc tcacagaact agtcagtttc agctgattca agaggagctg 1740
ctagagaaag cttcaaactc cagcaaactg gaaagtgaaa tgacaaagaa atgttctcaa 1800
cttttaactc ttgagaaaca gctggaagaa aagatagttg cttattcctc tattgctgca 1860
aaaaatgcag aactagaaca ggagcttatg gaaaagaatg aaaagataag gagtctagaa 1920
accaatatta atacagagca tgagaaaatt tgtttagcct ttgaaaaagc aaagaaaatt 1980
cacttggaac agcataaaga aatggaaaag cagattgaaa gagttaggca actagattca 2040
gcattggaaa tttgtaagga agaacttgtc ttgcatttga atcaattgga aggaaataag 2100
gaaaagtttg aaaaacagtt aaagaagaaa tctgaagagg tatattgttt acagaaagag 2160
ctaaagataa aaaatcacag tcttcaagag acttctgagc aaaacgttat tctacagcat 2220
actcttcagc aacagcagca aatgttacaa caagagacaa ttagaaatgg agagctagaa 2280
gatactcaaa ctaaacttga aaaacaggtg tcaaaactgg aacaagaact tcaaaaacaa 2340
agggaaagtt cagctgaaaa gttgagaaaa atggaggaga aatgtgaatc agctgcacat 2400
gaagcagatt tgaaaaggca aaaagtgatt gagcttactg gcactgccag gcaagtaaag 2460
attgagatgg atcagtacaa agaagagctg tctaaaatgg aaaaggaaat aatgcaccta 2520
aaacgagatg gagaaaataa agcaatgcac ctctctcaat tagatatgat cttagatcag 2580
acaaagacag agctagaaaa gaaaacaaat gctgtaaagg agttagaaaa gttacagcac 2640
agtactgaaa ctgaactaac agaagccttg caaaaacggg aagtacttga gactgaacta 2700
caaaatgctc atggagaatt aaaaagtact ttaagacaac tccaggaatt gagagatgta 2760
ctacagaagg ctcaattatc attagaggaa aaatacacta ctataaagga tctcacagct 2820
gaacttagag aatgcaagat ggagattgaa gacaaaaagc aggagctcct tgaaatggat 2880
caggcactta aagagagaaa ttgggaacta aagcaaagag cagctcaggt tacacatttg 2940
gatatgacta ttcgtgagca cagaggagaa atggaacaaa aaataattaa attagaaggt 3000
actctggaga aatcagaatt ggaacttaaa gaatgtaaca aacagataga aagtctgaat 3060
gacaaattac aaaatgctaa agaacaggtt cgagaaaaag agtttataat gctacaaaat 3120
gaacaggaga taagtcaact gaaaaaagaa attgaaagaa cacaacaaag gatgaaagaa 3180
atggagagtg ttatgaaaga gcaagaacag tacattgcca ctcagtacaa ggaggccata 3240
gatttggggc aagaattgag gctgacccgg gagcaggtgc agaactctca tacagaattg 3300
gcagaggctc gtcatcagca agtccaagca cagagagaaa tagaaaggct ctctagtgaa 3360
ctggaggata tgaagcaact ctctaaagag aaagatgctc atggaaacca tttagctgaa 3420
gaactggggg cttctaaagt acgtgaagct catttagaag caagaatgca agcagaaatc 3480
aagaaattgt cagcagaagt agaatctctc aaagaagctt atcatatgga gatgatttca 3540
catcaagaga accatgcaaa gtggaagatt tctgctgact ctcaaaagtc ttctgttcag 3600
caactaaacg aacagttaga gaaggcaaaa ttggaattag aagaagctca ggatactgta 3660
agcaatttgc atcaacaagt ccaagatagg aatgaagtaa ttgaagctgc aaatgaagca 3720
ttacttacta aagaatcaga attaaccaga ttacaggcca aaatttctgg acatgaaaag 3780
gcagaagaca tcaagtttct gccagcccca tttacatctc caacagaaat tatgcctgat 3840
gttcaagatc caaaatttgc taaatgtttt cacacatctt tttccaagtg tacaaaatta 3900
cgtcgctcta ttagtgccag tgatcttact ttcaaaattc atggtgatga agatctttct 3960
gaagaattac tacaggactt aaagaaaatg caattagaac agccttcaac attagaagaa 4020
agccataaga atctgactta cacccagcca gactcattta aacctctcac atataaccta 4080

gaagctgata gttctgagaa taatgacttt aacacgctta gtgggatgct aagatacata 4140
aacaaagaag taagactatt aaaaaagtct tctatgcaaa caggtgctgg tttaaatcag 4200
ggagaaaatg tgtaattcaa agaagatact gatgtgttga aaaaatggaa tttttggtac 4260
tgtgctgttt acttattata tgtagctcat acttcataga agctgttatt ttgcttttga 4320
ataaatttta tatttcaata ttttaaaaga aagcccttct aaaacttaat tatattttta 4380
aagaaaattt aaaaaaaaaa aaaaaggggg 4410

<210> 26
<211> 5032
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1758089CB1

<400> 26
ccggcccgag cggggcctgg gggtgcgacg ccgagggcgg gggagagcgc gccgctgctc 60
ccggaccggg ccgcgcacgc cgcctcagga accatcactg ttgctggagg cacctgacaa 120
atcctagcga atttttggag catctccacc caggaacctc gccatccaga agtgtgcttc 180
ccgcacagct gcagccatgg ggtctgagga ccacggcgcc cagaacccca gctgtaaaat 240
catgacgttt cgcccaacca tggaagaatt taaagacttc aacaaatacg tggcctacat 300
agagtcgcag ggagcccacc gggcgggcct ggccaagatc atccccccga aggagtggaa 360
gccgcggcag acgtatgatg acatcgacga cgtggtgatc ccggcgccca tccagcaggt 420
ggtgacgggc cagtcgggcc tcttcacgca gtacaatatc cagaagaagg ccatgacagt 480
gggcgagtac cgccgcctgg ccaacagcga gaagtactgt accccgcggc accaggactt 540
tgacgacctt gaacgcaaat actggaagaa cctcaccttt gtctccccga tctacggggc 600
tgacatcagc ggctctttgt atgatgacga cgtggcccag tggaacatcg ggagcctccg 660
gaccatcctg gacatggtgg agcgcgagtg cggcaccatc atcgagggcg tgaacacgcc 720
ctacctgtac ttcggcatgt ggaagaccac cttcgcctgg cacaccgagg acatggacct 780
gtacagcatc aactacctgc actttgggga gcctaagtcc tggtacgcca tcccaccaga 840
gcacggcaag cgcctggagc ggctggccat cggcttcttc cccgggagct cgcagggctg 900
cgacgccttc ctgcggcata agatgaccct catctcgccc atcatcctga agaagtacgg 960
gatccccttc agccggatca cgcaggaggc cggggaattc atgatcacat ttccctacgg 1020
ctaccacgcc ggcttcaatc acgggttcaa ctgcgcagaa tctaccaact tcgccaccct 1080
gcggtggatt gactacggca aagtggccac tcagtgcacg tgccggaagg acatggtcaa 1140
gatctccatg gacgtgttcg tgcgcatcct gcagcccgag cgctacgagc tgtggaagca 1200
gggcaaggac ctcacggtgc tggaccacac gcggcccacg gcgctcacca gccccgagct 1260
gagctcctgg agtgcgtccc gggcctcgct gaaggccaag ctcctccgca ggtctcaccg 1320
gaaacggagc cagcccaaga agccgaagcc cgaagacccc aagttccctg ggagggtac 1380
ggctggggca gcgctcctag aggaggctgg gggcagcgtg aaggaggagg ctgggccgga 1440
ggttgacccc gaggaggagg aggaggagcc gcagccactg ccacacggcc gggaggccga 1500
gggcgcagaa gaggacggga ggggcaagct gcggccaacc aaggccaaga gcgagcggaa 1560
gaagaagagc ttcggcctgc tgcccccaca gctgccgccc ccgcctgctc acttcccctc 1620
agaggaggcg ctgtggctgc catccccact ggagcccccg gtgctgggcc caggccctgc 1680
agccatggag gagagccccc tgccggcacc ccttaatgtc gtgccccctg aggtgcccag 1740
tgaggagcta gaggccaagc ctcggcccat catccccatg ctgtacgtgg tgccgcggcc 1800
gggcaaggca gccttcaacc aggagcacgt gtcctgccag caggcctttg agcactttgc 1860
ccagaagggt ccgacctgga aggaaccagt ttcccccatg gagctgacgg ggccagagga 1920
cggtgcagcc agcagtgggg caggtcgcat ggagaccaaa gccccgggccg gagaggggca 1980
ggcaccgtcc acattttcca aattgaagat ggagatcaag aagagccggc gccatcccct 2040
gggccggccg cccacccggt ccccactgtc ggtggtgaag caggaggcct caagtgacga 2100
ggaggcatcc cctttctccg gggaggaaga tgtgagtgac ccggacgcct tgaggccgct 2160
gctgtctctg cagtggaaga acagggcggc cagcttccag gccgagagga agttcaacgc 2220
agcggctgcg cgcacggagc cctactgcgc catctgcacg ctcttctacc cctactgcca 2280
ggccctacag actgagaagg aggcacccat agcctccctc ggagagggct gcccggccac 2340

attaccctcc aaaagccgtc agaagacccg accgctcatc cctgagatgt gcttcacctc 2400
tggcggtgag aacacggagc cgctgcctgc caactcctac atcggcgacg acgggaccag 2460
ccccctgatc gcctgcggca agtgctgcct gcaggtccat gccagttgct atggcatccg 2520
tcccgagctg gtcaatgaag gctggacgtg ttcccggtgc gcggcccacg cctggactgc 2580
ggagtgttgc ctgtgcaacc tgcgaggagg tgcgctgcag atgaccaccg ataggaggtg 2640
gatccacgtg atctgtgcca tcgcagtccc cgaggcgcgc ttcctgaacg tgattgagcg 2700
ccaccctgtg gacatcagcg ccatccccga gcagcggtgg aagctgaaat gcgtgtactg 2760
ccggaagcgg atgaagaagg tgtcaggtgc ctgtatccag tgctcctacg agcactgctc 2820
cacgtccttc cacgtgacct gcgcccacgc cgcaggcgtg ctcatggagc cggacgactg 2880
gccctatgtg gtctccatca cctgcctcaa gcacaagtcg gggggtcacg ctgtccaact 2940
cctgagggcc gtgtccctag gccaggtggt catcaccaag aaccgcaacg ggctgtacta 3000
ccgctgtcgc gtcatcggtg ccgcctcgca gacctgctac gaagtgaact tcgacgatgg 3060
ctcctacagc gacaacctgt accctgagag catcacgagt agggactgtg tccagctggg 3120
accccccttcc gagggggagc tggtggagct ccggtggact gacggcaacc tctacaaggc 3180
caagttcatc tcctccgtca ccagccacat ctaccaggtg gagtttgagg acgggtccca 3240
gctgacggtg aagcgtgggg acatcttcac cctggaggag gagctgccca agagggtccg 3300
ctctcggctg tcactgagca cgggggcacc gcaggagccc gccttctcgg gggaggaggc 3360
caaggccgcc aagcgcccgc gtgtgggcac ccgcttgcc acggaggact ccgggcggag 3420
ccaggactac gtggccttcg tggagagcct cctgcaggtg cagggccggc ccggagcccc 3480
cttctaggac agctggccgc tcaggcgacc ctcagcccgg cggggaggcc atggcatgcc 3540
ccgggcgttc gcttgctgtg aattcctgtc ctcgtgtccc cgacccccga gaggccacct 3600
ccaagccgcg ggtgcccccct agggcgacag gagccagcgg gacgccgcac gcggccccag 3660
actcagggag cagggccagg cgggctcggg ggccggccag gggagcaccc cactcaacta 3720
ctcagaattt taaaccatgt aagctctctt cttctcgaaa aggtgctact gcaatgccct 3780
actgagcaac ctttgagatt gtcacttctg tacataaacc acctttgtga ggctctttct 3840
ataaatacat attgtttaaa aaaaagcaag aaaaaaagga aaacaaagga aaatatcccc 3900
aaagttgttt tctagatttg tggctttaag aaaaacaaaa caaaacaaac acattgtttt 3960
tctcagaacc aggattctct gagaggtcag agcatctcgc tgtttttttg ttgttgtttt 4020
aaaatattat gatttggcta cagaccaggc agggaaagag acccggtaat tggagggtga 4080
gcctcggggg gggggcagga cgcccccggtt tcggcacagc ccggtcactc acggcctcgc 4140
tctcgcctca ccccggctcc tgggctttga tggtctggtg ccagtgcctg tgcccactct 4200
gtgcctgctg ggaggaggcc caggctctct ggtggccgcc cctgtgcacc tggccagggg 4260
aagcccgggg gtctggggcc tccctccgtc tgcgcccacc tttgcagaat aaactctctc 4320
ctggggtttg tctatctttg tttctctcac ccgagagaaa cgcaggtgtt ccagaggctt 4380
ccttgcagac aaagcacccc tgcacctccc atggctcagg atgagggagg cccccaggcc 4440
cttctggttg gtagtgagtg tggacagctt cccagctctt cgggtacaac cctgagcagg 4500
tcgggggaca cagggccgag gcaggccttc ggggcccctt tcgcctgctt ccgggcaggg 4560
acgaggcctg gtgtcctcgc tccacccacc cacgctgctg tcacctgagg ggaatctgct 4620
tcttaggagt gggttgagct gatagagaaa aaacggcctt cagcccaggc tgggaagcgc 4680
cttctccagg tgcctctccc tcaccagctc tgcacccctc tggggagcct tccccacctt 4740
agctgtctcc tgccccaggg agggatggag gagataattt gcttatatta aaaacaaaaa 4800
atggctgagg caggagtttg ggaccagcct gggctatata gcaagacccc atcactacaa 4860
attttttaca aattagctag gtgtggtggt gcgcacctgt ggtcccagct actcgggagg 4920
ctgtggtggg aggattgctt gagtccagga ggttgaggct gcagtcagct cagattgcac 4980
cactgcactc cagcctgggc aacagagcga gaccctgtct ccaaaaaaaa aa 5032

<210> 27
<211> 1355
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3533891CB1

<400> 27

cggggacgcg cgcccggcct gtcgctgtgg aaaccgctag gccagcgctc gccgggacct 60
ggaatccctg tacgccgagg tgggagccgg tggaccggtc cccagccgg cccccacctc 120
cgcttcccgg tgtttgaggg ttcgggcctc ccgccgggga gttcacccct cgggctcgtc 180
agtagggctg tggctgtcgc ctcttcctgc agcgccaggc tccgcccggt ctcacagtcg 240
gcttaggggc tttgcgtgca ctgcggttgg gtggaaaaac ccactcctgg ttgtttagac 300
gttggcctgc agacgatgtc atttctgtat tcctctaagg caggaagtca ttatgcaact 360
tacacatatt catcagattt cctctgactt acccggacat gtacatggga atgatgtgca 420
ctgccaagaa atgtgggatt aggtttcagc ctccagctat tatcttaatc tatgagagtg 480
aaatcaaggg gaaaattcgc cagcgcatta tgccagttcg aaacttttca aagttttcag 540
attgcaccag agctgctgaa caattaaaga ataatccgcg acacaagagt tacctagaac 600
aagtatccct gaggcagcta gagaagctat tcagtttttt acgaggttac ttgtcggggc 660
agagtctggc agaaacaatg gaacaaattc aacgggaaac aaccattgat cctgaggaag 720
acctgaacaa actagatgac aaggagcttg ccaaaagaaa gagcatcatg gatgaacttt 780
ttgagaaaaa tcagaagaag aaggatgatc caaattttgt ttatgacatt gaggttgaat 840
ttccacagga cgatcaactg cagtcctgtg gctgggacac agagtcagct gatgagttct 900
gataccaaac actcaaaaca tgcattgggc tagcagaata tccatgttta ttaccagact 960
ggttctggaa gaagctgtaa agaatactaa atatgttggg ttatagggga ttgaccatgt 1020
tacttttcaa aaccaggaca tttaaagcat ctactatgta ggtgcatgag gagtatggga 1080
aaaacagaat aaaggaatct gcctttaagg agcttacaat catgccgggt gcggtggctc 1140
acgcctgtaa tcccagcact ttgggaggct gaggcgggtg gatcacctaa ggtcaggagt 1200
tcgagaccag cctagccaac atggtgaaac ctcgcctcta ctaaaaatac aaaaattagc 1260
caggcgtggt ggcgggtgcc tgtaatcccg gctactcagg aggctgaggc aggagattcg 1320
cttgaacctg ggattaactg acgttgcagt gagcc 1355

<210> 28
<211> 4912
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1510943CB1

<400> 28
cggccccag cggcggcagc ggagagcgcg gtcccgggtc ggagcctggg acacctccgc 60
acggacgggg cgggcggcgc ggacaggcca tggggacccg ggccgggcca gcggtggcgg 120
gccagcggga gccccggcc tgagaagtgg gcggcggggt ggcgggggcc atgacctcgg 180
tgtggaagcg cctgcagcgc gtgggcaagc gggcggccaa gttccagttc gtggcctgtt 240
accacgagct agtgttggag tgcaccaaga aatggcagcc agataagctg gtggtggtat 300
ggacccgtcg gaaccgacgc atctgctcca aggcccacag ctggcagccg ggcatccaga 360
acccataccg ggcaccgtg gtgtggatgg tacctgagaa tgtggacatc tctgtgaccc 420
tctacaggga cccccacgtg gaccagtatg aggccaaaga gtggacattt attattgaaa 480
atgagtctaa ggggcagcgg aaggtgctgg ccacggccga ggtggacctg gcccgccatg 540
cagggcccgt gcctgtccaa gtcccactga ggctgcggct gaagccaaag tcagtgaagg 600
tggtgcaggc tgagctgagc ctcactcttt ccggggtgct gctgcgggag ggccgtgcca 660
cggacgatga catgcagagt ctcgcaagcc tcatgagtgt gaagcctagt gatgtgggca 720
acttggatga ctttgctgag agtgatgaag atgaggctca tggcccagga gccccggagg 780
cccgggctcg agtcccccag ccagatccct ctcgagagct gaagacgctt tgtgaggagg 840
aggaggaagg ccaaggacga ccccagcagg cagttgccag cccttctaat gctgaggata 900
ccagcccagc ccctgtgagt gctcctgcac ccccagccag aacctcccga ggccaggggt 960
cagaacgagc taatgaagcg gggggccagg taggccctga ggccccaagg cccccggaaa 1020
cctcaccaga gatgaggtct tcaaggcagc cagcccagga cacggccccc accccagccc 1080
ctcggctccg gaaaggctct gatgccctcc ggcccccagt cccccagggg gaagatgagg 1140
tccccaaagc ctcaggggct cctccagcag gattgggctc tgctagggag acccaggccc 1200
aggcatgccc tcaggaaggg acagaagccc atggagctag gctgggcccg agcattgagg 1260
ataaaggttc tggagaccct tttggaaggc agagactcaa ggctgaagag atggacactg 1320

aggacaggcc agaggccagt ggggtggaca ctgagccaag gtcaggaggc agagaggcaa 1380
acactaagag gtcaggagtc agagctgggg aggctgaaga gagttcagca gtttgtcaag 1440
tggatgctga gcagaggtca aaggtgagac atgtggacac taagggacca gaggcgacag 1500
gggtgatgcc tgaggcaaga tgcaggggga cccctgaggc tcctccaagg ggctctcagg 1560
ggaggctggg agtcaggacc agggatgagg ctccctcagg cctgagcctg cccccagcgg 1620
agcctgcagg gcactctggg caacttggtg acctcgaggg ggccagggct gctgcaggcc 1680
aggagagaga gggtgcagaa gtgaggggtg gagcacctgg tattgagggg acaggcctgg 1740
agcagggccc ttctgttgga gcaataagca ccaggcccca ggtgagcagc tggcaggggg 1800
ccctgttatc aactgcccag gggcaatat ccaggggtct gggaggctgg gaggcagaag 1860
ctgggggttc aggggtcctg gaaacagaga ctgaggtggt agggttggag gtgctgggaa 1920
cccaggagaa agaagttgag gggtcagggt tcccagagac taggacacta gaaattgaga 1980
tattgggggc cttggagaaa gaagcagcaa gatcaagggt cctggagtca gaggttgctg 2040
ggacagcaca gtgtgaggga ctggagaccc aggaaacaga ggtgggggtc atagagaccc 2100
cagggacaga gactgaggta ttggggaccc agaaaacaga agctgggggt tcaggagttt 2160
tgcagacaag aactacgata gcagagactg aggtactggt gacccaggag atatctgggg 2220
atttagggcc actgaagata gaagatacaa tacagtctga gatgctgggg acccaggaga 2280
cagaggtgga agcttctagg gtaccagagt cagaggctga ggggacagaa gctaaaatat 2340
tagggaccca ggagataaca gctagggatt caggggtcag agagatagaa gcagagatag 2400
cagagtctga catattggta gcccaggaga tagaggtggg acttttgggg gttctgggaa 2460
tagagactgg ggcagcagaa ggtgcgatat tggggaccca agagatagca tctagggatt 2520
caggggtccc agggttagaa gctgatacaa cagggatcca ggtgaaagag gttggggtt 2580
cagaggttcc agagatagcg actgggacag cagaaactga gatattgggg acccaagaga 2640
tagcatctag gagttcaggg gtcccagggc tagaatctga ggtagctggg gcccaggaga 2700
cagaggtcgg gggttcaggg atctcagggc ccgaggctgg aatggcagag gcccgagtac 2760
tgatgacccg taagacagaa attatagttc cagaggctga gaaggaagag gctcagactt 2820
cgggggtcca ggaagcagag actagagttg ggagtgctct caaatatgag gctttaaggg 2880
ccccagtcac tcagccaaga gttttaggat cccaggaagc aaaagcagag atttcaggag 2940
tacaagggtc agagactcaa gttctgagag tccaggaggc agaggctggg gtttgggggga 3000
tgtcagaggg caaatctggg gcttgggggg cccaggaagc agagatgaag gttttagagt 3060
ctccagagaa caaatctggt acttttaagg cccaggaagc ggaggctggg gtcttgggaa 3120
atgagaaggg gaaagaagct gagggaagcc tcacagaggc cagcctgcct gaagcacagg 3180
tggccagtgg ggcaggggct ggggcgccca gggcctcttc cccagagaag gctgaagagg 3240
acaggaggct gccgggcagc caggcaccac ctgccctggt cagctccagc cagtccctgc 3300
tggagtggtg ccaggaagtc accactggct accgtggcgt ccgcatcacc aacttcacca 3360
catcctggcg caacggcttg gccttctgtg ccatcctgca ccgattctac ccagacaaga 3420
ttgactatgc ctcgctagac ccactcaaca tcaagcagaa caacaagcag gccttcgatg 3480
gcttcgcggc tctgggcgtg tcgcggctgc tggagcccgc ggacatggtg ctactgtcgg 3540
tgcccgacaa gctcatcgtc atgacgtacc tgtgccagat ccgcgccttc tgcaccgggc 3600
aggagctgca gctggtacaa ctggagggcg gcggcggcgc cggcacgtac cgcgtgggca 3660
gcgcccagcc cagcccgccc gacgacctgg acgccggagg cctggcgcag cggctgcgcg 3720
gtcacggggc cgaggggccc caggagccca aggaggccgc agaccgcgca gacggggcgg 3780
ccccgggggt ggcctccagg aacgcggtcg cggccgcgc ctccaaggac ggcggggccg 3840
aggcccccg agagtcgcga cccgcggagg tcccggccga ggggctggtg aacggggcgg 3900
gggcaccggg cggcggcggc gtgaggctgc gacggccctc ggtcaacggg gagcccgggt 3960
cggtgccccc gccccgcgcg cacggctcct tctcccacgt gcgcgacgcg gacctgctca 4020
agaagaggcg ctcgcggctg cggaacagca gctcgttctc gatggacgat ccggacgcgg 4080
gagccatggg agctgcggct gcagaaggcc aggcccctga ccccagccct gccccaggcc 4140
cacccacagc tgcagactct caacagcccc ctggtgggag ttccccctcg gaggaaccac 4200
ccccaagccc aggggaggag gctgggctgc aacggttcca ggacacaagt cagtacgtgt 4260
gtgcagagct gcaggccctg gaacaggagc agaggcagat agatgggcgg gcggctgagg 4320
tggagatgca gctgaggagc ctcatggagt caggtgccaa caagctgcag gaggaggtgc 4380
tgatccagga gtggttcacc ctggtcaaca agaagaacgc tctcatccgg aggcaggacc 4440
agctgcagct gctcatggag gagcaggact tggagcgaag gttcgagctg ctgagccgcg 4500
agctgcgggc catgctggcc atcgaagact ggcagaaaac gtccgctcag cagcaccgag 4560
agcagctcct actggaggag ctggtgtcgc tggtgaacca gcgcgatgag ctagtccggg 4620
acctggacca caaggagcgg atcgccctgg aggaggacga gcgcctggag cgcggcctgg 4680

aacagcggcg ccgcaagctg agccggcagt tgagccggcg ggagcgctgc gtgctgagct 4740
gaggccgccg gcccgggtgg cccataactt ctcgcgtccc cggcgtccgc cgccgccccg 4800
ggcctgcgct gcggacgacc cggccgtccc ggaggccgcg cgcgtgtccg ctagggccg 4860
ccggcgccct tccccgtata gggcagggcg gatccccgac cccacgggcg gg 4912

<210> 29
<211> 2241
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2119377CB1

<400> 29
cccacgcgtc cgcgaggtag cggtggcctg cagcggcctc ctcccgcag tgaagcatgg 60
gccagaagct ctcggggagc ctcaagtcag tggaggtgcg agagccggcg ctgcggccgg 120
ccaagcggga gctgcggggt gcagagcccg ggcggccggc gcggctggac cagctgttgg 180
acatgccagc ggcggggctg gctgtgcagc tgcggcacgc gtggaacccc gaggaccgct 240
cgctcaacgt cttcgtcaag gacgacgacc ggctcacctt ccaccggcac cccgtggccc 300
agagcaccga cggcatccgc ggcaaggtgg gccacgcccg cggcctgcac gcctggcaga 360
tcaactggcc ggctcggcag cgcggcaccc acgctgtagt tggtgtggcc acggcccgtg 420
ctcccctgca ctccgtgggc tacacggcgc tggtaggcag tgacgccgag tcgtggggct 480
gggacctggg ccgcagccgc ctctaccacg acggcaagaa ccagcccggc gtggcctacc 540
cggcctttct ggggcccgac gaggcctttg cgctgcccga ctcgctgctc gtggtgctgg 600
acatggatga gggcacactc agcttcatcg tggatggcca gtacctgggc gtggccttcc 660
gaggtctcaa gggcaagaag ctgtacccgg tggtgagtgc cgtgtggggc cactgtgaag 720
tcaccatgcg ctacatcaac ggccttgacc ccgagcccct gccactgatg gacctgtgcc 780
ggagatccat ccgctcggcc ctgggccgcc agcgcctgca ggacatcagc tccctgcccc 840
tgcctcagtc tctcaaaaac tatctgcagt accagtgagc caagcctgat gggcagcaca 900
gacacagaca cacaccgcag ggcccgaccc tcctgtcatt cacagtccca tggcacatag 960
gggaaaggat ctacccttct cctggctccc caggacactc agttctttca aagaccagga 1020
tgtggtacca actttggaaa cgaaaggtct cttgccaaca gtatctactg ccctcgaggc 1080
agccctccca agtcagacac ctccttcgga gccacagaga gcctggagtc tgcacctcct 1140
ggaaatcctg ccaccaacca ggacacagca gccaccgtat tgatcagaga gcctgtttcc 1200
ttattcaaga gaatgaataa aacatttagg caggagactt tctattgtgt gccccgttgc 1260
agacagggcc agggagaaat tagccagtgc agggggaaaa ttgcctctga ataatgaatg 1320
atgaaagcct gacgccgtgc ccctcctggc ccatacgcct tgccagggcg gcaggattgt 1380
cacaccgttt taggactctg tgccactttg agagactgtt cccaggaggc ccaaccgcag 1440
acctggcaag tggacagtgc agtgtggaga caccttccgg cttacctctt tgaacgttgt 1500
ttacctaccc ctttccacgt gctccccttc ccagccactg actcacagtt ctctggatgc 1560
ccagacacct ctcttcaggg aagatgagtc tgactggttt gccccaggga agagcgtatc 1620
cttaccatta ttttaagtag catttgcatt ttaaaagagg aatgcggaga ggacagtact 1680
tagtccaaag gtgctaacgg gggaactggg ggcattgtga cacccaagtc tgatgtgtcc 1740
tgggttgggg ggccttcctg gagtgtcagg gtctctgggc aacgtccatt cagggtgcgg 1800
catggctgtc acaaagcttt atttgagcaa attatttttt cactttagga gacttctaca 1860
agtttgtttc ctgtttcaaa tgtgtgtgtg atgtgctgtt tatttatcag cttgaggtcc 1920
atgggggcag ccttgtgact ggaagggtgg atatgggaga cacattctct acctgctccg 1980
agcctggtcc tctcgcagga atgctgctgc tgcctccgcc gccactgctg ctgccacctc 2040
ttatatgttt cagacactct ctgcccagac tcatttttaa ctggaaatca tcacagcagt 2100
gggatatcag agccccagac agcactgcct cttcctcccc accccactgc ccccacctta 2160
atgtgaattt gactgatgaa tgaagagcgt ttctaataaa gtttgtcatt cagtccttaa 2220
aaaaaaaaaa aaaaaaaaaa a 2241

<210> 30
<211> 1853

<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3176058CB1

<400> 30
gaacggacc gctttcccgg aagtgcttgc ggcctctgcc cagcgagctg ccccggggtc 60
tctctggttt cctaatcagg gcaacgccgc gggagagaac ctttaccttg gctgcactaa 120
gttctcggtg ccactccctg gcagggcggg accttgttta ggccctgtga tcgcgcggtt 180
cgtagtagcg caaggcgcag agtggacctt gacccgccta gggcgggaag agtttggccc 240
gccgggtccc aaagggcaga atggacgggc tcctaaatcc cagggaatcc tctaaattca 300
ttgcagaaaa cagtcgggat gtgtttattg acagcggagg cgtacggagg gtggcagagc 360
tgctgctggc caaggcggcg gggccagagc tgcgcgtgga ggggtggaaa gcccttcatg 420
agctgaaccc cagggcggcc gacgaggccg cggtcaactg ggtgttcgtg acagacacgc 480
tcaacttctc cttttggtcg gagcaggacg agcacaagtg tgtggtgagg tacagaggga 540
aaacatacag tgggtactgg tccctgtgcg ccgccgtcaa cagagccctc gacgaaggga 600
taccaataac tagtgcctcg tactacgcga cagtgaccct ggatcaggtt cggaatatac 660
ttcgttctga cacagacgtt tccatgcctt tagtagaaga gaggcatcgg attctcaatg 720
aaaccgggaa aattctgctg gagaagtttg gaggctcttt tctcaactgc gtccgagaaa 780
gtgagaatag tgcgcagaag ttaatgcacc tggtggttga aagttttcct tcttacagag 840
atgtgactct gtttgagggg aaaagagttt ctttttacaa acgagcccaa atccttgtag 900
cagatacgtg gagtgtattg gaaggaaaag gagatggctg cttcaaggac atctccagta 960
tcaccatgtt tgctgattat agattacctc aggttcttgc tcatcttgga gccctgaaat 1020
actctgatga cctactgaag aagcttctca aaggagaaat gctctcatat ggagacaggc 1080
aagaggtgga aatcagaggg tgctcgcttt ggtgtgttga gctgatccgg gattgtcttc 1140
tggagcttat tgaacaaaag ggtgaaaaac ctaatggaga gatcaattcc attcttctgg 1200
attattactt atgggactat gcccatgacc atagggaaga tatgaaagga attccgtttc 1260
atcgcatacg ttgcatatat tattgacctc aagtgtaaac tgatccaaag aaaacccccct 1320
gcgttttata tcatatcatc tgtacagttt tgctttgata tttagagaac atgatcgagg 1380
ttataggaaa ttgattgccc attctcactt gaaaaatact tcctaggccg ggcacagtgg 1440
tttatgcctg taatcccagc actgtgggag gctgaggcgg gtggatcatc tgaggtcagg 1500
agtttgagac cacctggcca acatggtgaa accccatctc tactaaaaat acaaaattag 1560
ctgggcgtgg tggcacgtgc ctgtaatccc agttacttgg gaggctgagg caagagaatc 1620
gcttgaactc agaaggtgga ggttgcagtg aggcgagatt aggccattgc actccagcct 1680
cagcaacaag agtgaaactt tgtctcaaaa aacaacaaca acaacaacaa caacaacaaa 1740
acttcctagg ccagatgtgg tggctcatat gtataatctt agcactttgg gaggccaagg 1800
caggaggatt gcttgaggcg aggagttcaa gaccagccta ggcaacatag gga 1853

<210> 31
<211> 2541
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2299818CB1

<400> 31
aaaacattct tggccaaaat ctaggggaag ttactgccac ttcgtactat ataaggaaaa 60
caaagacacc atggatgcta ttaatgtact ctccaaatac ttaagagtca agccaaatat 120
attctcctac atgggaacca aagataaaag ggctataaca gttcaagaaa ttgctgttct 180
caaaataact gcacaaagac ttgcccacct gaataagtgc ttgatgaact ttaagctagg 240
gaatttcagc tatcaaaaaa acccactgaa attgggagag cttcaaggaa accacttcac 300
tgttgttctc agaaatataa caggaactga tgaccaagta cagcaagcta tgaactctct 360

caaggagatt ggatttatta actactatgg aatgcaaaga tttggaacca cagctgtccc 420
tacgtatcag gttggaagag ctatactaca aaattcctgg acagaagtca tggatttaat 480
attgaaaccc cgctctggag ctgaaaaggg ctacttggtt aaatgcagag aagaatgggc 540
aaagaccaaa gacccaactg ctgccctcag aaaactacct gtcaaaaggt gtgtggaagg 600
gcagctgctt cgaggacttt caaaatatgg aatgaagaat atagtctctg catttggcat 660
aatacccaga aataatcgct taatgtatat tcatagctac caaagctatg tgtggaataa 720
catggtaagc aagaggatag aagactatgg actaaaacct gttccagggg acctcgttct 780
caaaggagcc acagccacct atattgagga agatgatgtt aataattact ctatccatga 840
tgtggtaatg cccttgcctg gtttcgatgt tatctaccca aagcataaaa ttcaagaagc 900
ctacagggaa atgctcacag ctgacaatct tgatattgac aacatgagac acaaaattcg 960
agattattcc ttgtcagggg cctaccgaaa gatcattatt cgtcctcaga atgttagctg 1020
ggaagtcgtt gcatatgatg atcccaaaat tccacttttc aacacagatg tggacaacct 1080
agaagggaag acaccaccag tttttgcttc tgaaggcaaa tacagggctc tgaaaatgga 1140
tttttctcta ccccccttcta cttacgccac catggccatt cgagaagtgc taaaaatgga 1200
taccagtatc aagaaccaga cgcagctgaa tacaacctgg cttcgctgag cagtaccttg 1260
tccacagatt agaaaacgta cacaagtgtt tgcttcctgg ctccctgtgc atttttgtct 1320
tagttcagac tcatatatgg atttcaaatc tttgtaataa aaattatttg tatttttaag 1380
tttttattag cttaaagaaa taatttgcaa tatttgtaca tgtacacaaa tcctgaggtt 1440
cttaatttta gctcagaata taaattagtc aaaatacact tcaggtgctt aaatcagagt 1500
aaaatgtcag ctttacaata ataaaaaaag gactttggtt taaagtagca ggtttaggtt 1560
ttgctacatt ctcaaaagac agcaggagta tttgacacat ctgtgatgga gtatacaaca 1620
atgcatttta agagcaaatg caacaaaaca aatctggact atggataaat aatttgagag 1680
ctgccaccca caaatataaa tacagtactc atgctgactg aaataataag acatctacaa 1740
atttataaac aaaaagtgat tgtcattatc ctgcttatgt actagattca ggcaagcatt 1800
atagactttt tggttgcggt ggcttttgca tttatattat caatgccttg caggaacgtt 1860
gcattgatag gcccatttta tttttttatt ttttttttcg agacaggatc tcactctgta 1920
gcacaggctg gattgcagtg caatcctgca attctcaatc ttgcactgca gcctcgacct 1980
cccaggctcc agtgactctc ccacctcagc ctcctaagta gctgggagta caggcgcgca 2040
ccaccacgcc tagctgattt ttgtattttt ttgtagagac gggggtttgg ccatgttgcc 2100
gaggctaact cctgggatta caggcatgag ctgtgctggc cgggtttttt tttcttgatg 2160
taaacgtgta cagctgtttt attagttaag gtctaatttt tactctaggt gccttttatg 2220
ttcagaactc tttccactgg actggtattt gctcaaaaat aaataatggt agagaagaaa 2280
actataaaaa tggacaaggc tttcttctat cagtagcgtt taccctttgt caccagtggc 2340
tttggtattt ccatgtctgg cattgcataa acttctctgg tgtgaaagga taaatatgcc 2400
tttctaaagt tgtatatcaa aattgtatca atttttattt tctatgattt ctagaaacaa 2460
atgtaataaa tatttttaaa atctcctttc tactggttat gtaaataaat caaataaata 2520
tatcaaaaaa aaaaaaaaaa a 2541

<210> 32
<211> 4144
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2729451CB1

<400> 32
gtcgagatgg agcccaactc actccagtgg gtcggctcac cgtgtggctt gcacggacct 60
tacattttct acaaggcttt tcaattccac cttgaaggca aaccaagaat tttgtccctt 120
ggcgactttt tctttgtaag atgtacgcca aaggatccga tttgcatagc ggagctccag 180
ctgttgtggg aagagaggac cagccggcaa cttttatcca gctctaaact ttatttcctc 240
ccagaagaca ctcccccaggg cagaaatagc gaccatggcg aggatgaagt cattgctgtt 300
tccgaaaagg tgattgtgaa gcttgaagac ctggtcaagt gggtacattc tgatttctcc 360
aagtggagat gtggcttcca cgctggacca gtgaaaactg aggccttggg aaggaatgga 420
cagaaggaag ctctgctgaa gtacaggcag tcaaccctaa acagtggact caacttcaaa 480

gacgttctca aggagaaggc agacctgggg gaggacgagg aagaaacgaa cgtgatagtt 540
ctcagctacc cccagtactg ccggtaccgc tcgatgctga aacgcatcca ggataagcca 600
tcttccattc taacggacca gtttgcattg gccctggggg gcattgcagt ggtcagcagg 660
aaccctcaga tcctgtactg tcgggacacc tttgaccacc cgactctcat agaaaacgag 720
agtatatgcg atgagtttgc gccaaatctt aaaggcagac cacgcaaaaa gaaaccatgc 780
ccacaaagaa gagattcatt cagtggtgtt aaggattcca acaacaattc cgatggcaaa 840
gccgttgcca aggtgaaatg tgaggccagg tcagccttga ccaagccgaa gaataaccat 900
aactgtaaaa aagtctcaaa tgaagaaaaa ccaaaggttg ccattggtga agagtgcagg 960
gcagatgaac aagccttctt ggtggcactt tataaataca tgaaagaaag gaaaacgccg 1020
atagaacgaa taccctattt aggttttaaa cagattaacc tttggactat gtttcaagct 1080
gctcaaaaac tgggaggata tgaaacaata acagcccgcc gtcagtggaa acatatttat 1140
gatgaattag gcggtaatcc tgggagcacc agcgctgcca cttgtacccg cagacattat 1200
gaaagattaa tcctaccata tgaaagattt attaaaggag aagaagataa gcccctgcct 1260
ccaatcaaac ctcggaaaca ggagaacagt tcacaggaaa atgagaacaa aacaaaagta 1320
tctggaacca aacgcatcaa acatgaaata cctaaaagca agaaagaaaa agaaaatgcc 1380
ccaaagcccc aggatgcagc agaggtttca tcagagcaag aaaaagaaca agagacttta 1440
ataagccaga aaagcatccc tgagcctctc ccagcagcag acatgaagaa aaaaatagaa 1500
gggtatcagg aattttcagc gaagcccctg gcatccagag tagacccaga gaaggacaac 1560
gaaacagacc aaggttcaca cagtgagaag gtggcagagg aggcgggaga gaagggggcc 1620
acacctccac tcccaagtgc tcctctggcc ccagaaaaag attcagcctt ggtccctggg 1680
gccagcaaac agccactcac ctctcctagt gccctggtgg actcaaaaca agaatccaaa 1740
ctgtgctgtt ttacagagag ccctgaaagt gaaccccaag aagcatcctt ccccaccaca 1800
cagccacccgc tggcaaacca gaatgagacg gaggatgaca aactgcccgc catggcagat 1860
tacattgcca actgcaccgt gaaggtggac cagctgggca gtgacgacat ccacaatgcg 1920
ctcaagcaga ccccaaaggt ccttgtggtc cagtcgtttg acatgttcaa agacaaagac 1980
ctgactgggc ccatgaacga gaaccatgga cttaattaca cgcccctgct ctactctagg 2040
ggcaacccag gcatcatgtc cccactggcc aagaaaaagc ttttgtccca agtgagtggg 2100
gccagcctct ccagcagcta cccttatggc tccccacccc ctttgatcag caaaaagaaa 2160
ctgattgcta gggatgactt gtgttccagt ttgtcccaga cccaccatgg ccaaagcact 2220
gaccatatgg cggtcagccg gccatcagtg attcagcacg tccagagttt cagaagcaag 2280
ccctcggaag agagaaagac catcaatgac atctttaagc atgagaaact gagtcgatca 2340
gatccccacc gctgcagctt ctccaagcat caccttaacc cccttgctga ctcctacgtc 2400
ctgaagcaag aaattcagga gggcaaggat aaactcttag agaaaagggc cctcccccat 2460
tcccacatgc ctagcttcct ggctgacttc tactcgtccc ctcatctcca tagcctctac 2520
agacacaccg agcaccatct tcataatgaa cagacatcca aataccccttc cagggacatg 2580
tacagggaat cggaaaacag ttcttttcct tcccacagac accaagaaaa gctccatgta 2640
aattatctca cgtccctgca cctgcaagac aaaaagtcgg cggcagcaga agcccctacg 2700
gatgatcagc ctacagatct gagccttccc aagaacccgc acaaacctac cggcaaggtc 2760
ctgggcctgg ctcattccac cacagggccc caggagagca aaggcatctc ccagttccag 2820
gtcttaggca gccagagtcg agactgtcac cccaaagcct gtcgggtatc acccatgacc 2880
atgtcaggcc ctaaaaaata ccctgaatcg ctttcaagat caggaaaacc tcaccatgtg 2940
agactggaga atttcaggaa gatggaaggc atggtccacc caatcctgca ccggaaaatg 3000
agcccgcaga acattggggc ggcgcggccg atcaagcgca gcctggagga tttggacctt 3060
gtgattgcag ggaaaaaggc ccgggcagtg tctcccttag acccatccaa ggaggtctct 3120
gggaaggaga aggcctctga gcaggagagt gaaggcagca aagcagcgca cggtgggcat 3180
tccgggggcg gatcagaagg ccacaagctt cccctctcct cccctatctt cccaggtctg 3240
tattccggga gcctgtgtaa ctcgggcctc aactccaggc tcccggctgg gtattctcat 3300
tctctgcagt acttgaaaaa ccagactgtg ctttctccac tcatgcagcc cctggctttc 3360
cactcgcttg tgatgcaaag aggaattttt acatcaccga caaattctca gcagctgtac 3420
agacacttgg ctgcggctac acctgtagga agttcatatg gggacctttt gcataacagc 3480
atttaccctt tagctgctat aaatcctcaa gctgcctttc catcttccca gctgtcatcc 3540
gtgcacccca gtacaaaact gtaggctcag ctctgcccag cagtccaaag cggcatggcc 3600
aacagagctt cactccttac ccaggagtgc tggcttatag agttagaagt cagtatttct 3660
tctaatctga ggctatgatc agtcccagct gtaggggccc agaggggagg tgaacatgcc 3720
tgatttttgt gggacaactg tagcccacaa actgactggc tggtgagtct tgactccctt 3780
ccaacacaga tgcccaggca cctccagatc attcacttcg cacgtgggcc ttgtgaaggg 3840

atttgtgaat atccaggaag aacttagagg accccatctg agttcggatg gtcaggaaac 3900
aatctgggca aaaaagaggc aggcatttca aaggaagggg caaggaagac tggcaaacag 3960
atggcaaggg atgcccctct ttttcataaa actctccaag gttcaatcaa tgcaatgtat 4020
agtgaaactt caatagatct ttcattttga cactattaaa caatccagag aagtaaacac 4080
tgttaaattg actgtatata tttgcttctt aaaactacct gtatcactgt ttgctcacct 4140
aatt 4144

<210> 33
<211> 5218
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 878534CB1

<400> 33
attcgcgcgc gccttcccta gccacccggg gttgcctcct aacatggaat ggccaaagga 60
gcgccccttg cgggaagtga gggtgggttg ggactgggtc cgcgttgggg gaggtgcaat 120
cttcgggttt cgcctctcgg ctccctctgg ctctggagtt gggacccctt gtgggctctg 180
gaagtccgcc tgagacttgg gtcaaggagt caaactgtcg ccccccgctc ctcccccaga 240
aatccggtga gcggtaagga aagtgatgcc aagtcttcga agcctcagtg acaaacgcat 300
agcaagaaca catccactcc agagatacct tctcgaaaca aaagattttc ctacctgctt 360
atacttggta accgagggaa ttactaagac ttcttgctca tttctgagta ttgtctttat 420
atcctgacac tatgaatgct acttggatgc ctcttaagtc tgttctctgg ggaggcagta 480
aggggccgtg gagctggcct cggcctcggc atcgggagag gctggacttc ctgtctctct 540
gtgctgaatg gctgcgatgg cgcccgctct cactgacgca gcagctgaag cacaccatat 600
ccggttcaaa ctggctcccc catcctctac cttgtcccct ggcagtgccg aaaataacgg 660
caacgccaac atccttattg ctgccaacgg aaccaaaaga aaagccattg ctgcagagga 720
tcccagccta gatttccgaa ataatcctac caaggaagac ttgggaaagc tgcaaccact 780
ggtggcatct tatctctgct ctgatgtaac atctgttccc tcaaaggagt ctttgaagtt 840
gcaaggggtc ttcagcaagc agacagtcct taaatctcat cctctcttat ctcagtccta 900
tgaactccga gctgagctgt tggggagaca gccagttttg gagttttcct tagaaaatct 960
tagaaccatg aatacgagtg gtcagacagc tctgccacaa gcacctgtaa atgggttggc 1020
taagaaattg actaaaagtt caacacattc tgatcatgac aattccactt ccctcaatgg 1080
gggaaaacgg gctctcactt catctgctct tcatgggggt gaaatgggag gatctgaatc 1140
tggggacttg aagggggta tgaccaattg cactcttcca catagaagcc ttgatgtaga 1200
acacacaatt ttgtatagca ataatagcac tgcaaacaaa tcctctgtca attccatgga 1260
acagccggca cttcaaggaa gcagtagatt atcacctggt acagactcca gctctaactt 1320
gggggtgtc aaattggagg gtaaaaagtc tcccctgtct tccattcttt tcagtgcttt 1380
agattctgac acaaggataa cagctttact gcggcgacag gctgacattg agagccgtgc 1440
ccgcagatta caaaagcgct tacaggttgt gcaagccaag caggttgaga ggcatataca 1500
acatcagctg ggtggatttt tggagaagac tttgagcaaa ctgccaaact tggaatcctt 1560
gagaccacgg agccagttga tgctgactcg aaaggctgaa gctgccttga gaaaagctgc 1620
cagtgagacc accacttcag agggacttag caactttctg aaaagcaatt caatttcaga 1680
agaattggag agatttacag ctagtggcat agccaacttg aggtgcagtg aacaggcatt 1740
tgattcagat gtcactgaca gtagttcagg aggggagtct gatattgaag aggaagaact 1800
gaccagagct gatcccgagc agcgtcatgt acccctgaga cgcaggtcag aatggaaatg 1860
ggctgcagac cgggcagcta ttgtcagccg ctggaactgg cttcaggctc atgtttctga 1920
cttggaatat cgaattcgtc agcaaacaga catttacaaa cagatacgtg ctaataaggg 1980
gttgatagtt cttggggagg tacctccccc agagcataca acagacttat ttcttccact 2040
tagttctgag gtgaagacag atcatgggac tgataaattg attgagtctg tttctcagcc 2100
attggaaaac catggtgccc ctattattgg tcatatttca gagtcactgt ctaccaaatc 2160
atgtggagca ctcagacctg tcaatggagt tattaacact cttcagcctg tcttggcaga 2220
ccacattcca ggtgacagct ctgatgctga ggaacaatta cataagaagc aacgactgaa 2280
tctcgtctct tcatcatctg atggcacctg tgtggcagcc cggacacgtc ctgtactgag 2340

ctgtaagaag cggaggcttg ttcgacccaa cagcatcgtt cctctttcca agaaggttca 2400
ccggaacagc acaatccgcc ctggctgtga tgtgaatccc tcctgcgcac tgtgtggttc 2460
aggcagcatc aacaccatgc ctcccgaaat tcactatgaa gcccctctgt tggaacgtct 2520
ttcccagttg gactcttgtg ttcatcctgt tctagcattt ccagatgatg ttcccacaag 2580
cctgcatttc cagagcatgc tgaaatctca gtggcagaac aagccttttg acaaaatcaa 2640
acctcccaaa aagttatcgc ttaagcacag agcacccatg ccgggcagtc tgccagattc 2700
agctcgtaag gacaggcaca aattggtcag ctccttccta acaacagcca tgttgaagca 2760
tcacacagac atgagcagtt cgagctactt ggcagccacc caccatcctc cacacagtcc 2820
cttggtgcga cagctctcca cctcctcaga ttcccctgca cccgccagct ctagctcaca 2880
ggttacagcc agcacatcgc agcagccagt aaggaggaga aggggagaga gctcatttga 2940
tattaacaac attgtcatcc caatgtctgt tgctgcaaca actcgcgtag agaaactgca 3000
atacaaggaa atccttacgc ccagctggcg ggaggttgat cttcagtctc tgaaggggag 3060
tcctgatgag gagaatgaag agattgagga cctatccgac gcagccttcg ccgccctgca 3120
tgccaaatgt gaggagatgg agagggcacg gtggctgtgg accacgagtg tgccacccca 3180
gcggcggggc agcaggtcct acaggtcatc agacggccgg acaacccccc agctgggcag 3240
tgccaacccc tccacccccc agcctgcctc ccctgatgtc agcagtagcc actctttgtc 3300
agaatactcc catggtcagt cccctaggag ccccattagc ccggaactgc actcagcacc 3360
cctcacccct gtggctcggg acactctgcg acacttagcc agtgaggata cccgttgttc 3420
cacaccagag ctggggctgg atgaacagtc tgtccagccc tgggagcggc ggaccttccc 3480
cctggcgcac agtccccagg cggagtgtga ggaccagctg gatgcacagg agcgagcagc 3540
ccgctgcact cgacgcacct caggcagcaa gactggccgg gagacagagg cagcgcccac 3600
ctcgcctccc attgtccccc tcaagagtcg gcatctggtg gcagcagcca cagctcagcg 3660
cccgactcac agatgagcgg gagacagcca tctaaacaga ctcactaact attggcatta 3720
aagcttcaga aatctctgcg tttgatattc aaacatcata tgccggaaat tttcacagtt 3780
tttagtgaac ttaaggaatt tagatcctac tttggtattt tttttcttg ttttaatttt 3840
tgttttgttt ttgtttccat gttttcttgt cacacacctg agcacttcct cccgttggca 3900
aacagaagtt caggatgaga ccctgctggc ctggtcctgg cacatcctct gcactgttga 3960
atcactggac ttactgatct tagatgacca ccccctccct cacacctgtg ggcagggcag 4020
aacagcctgg cgggctacag tttagcatgg ccttcttgag ctagggtgga atggggcagg 4080
gtgctctgga ctcttacccc ctcccctccc atctgtggct tggctctgct gtggccctcc 4140
tggctgggtc cccttggttt ttcgtgctgg aacatcccca ccagagcctc tctgccataa 4200
ctgccagctg ctctccccga gtgctcagct ggcagaacac ctttccttct cacccagaac 4260
ttaagagact gattttttgt ttcatctgca tttggtcttc tctgttttga ctctttcact 4320
gcagtaacct ggctgtggct gctcaggttc ccctcctcat gcccccttggt acccttccct 4380
gtctgctctc ccatgccatg tacacaccca caacccgtcc ttccacttgg aatattttta 4440
ccacctatcc tgatctttga aggtagggtt aggactactt aacctctatt cccactcccc 4500
tgcaaactgg gggttgtggg aagtgagcag ccatctccct gtgtgatttt ttttttttt 4560
ccctctgatt cactttgcca tgtttccttc acatccagat ccctgtcggt gttagttcca 4620
ctcttggtct ttcacgctcc ccttgcctgt ggaacattgt ctggtcctag ctgtggttcc 4680
cattgttccc ccttcaccct tctctgttaa ccttgtgcct gtctcctgta tgatcacatc 4740
accaaaaagg gggagggggg agaagactct tttttttgg ccattttgta atcgtataaa 4800
aatagtagac aactgcttaa tggttggggt tttttcacaa ttttcaacat tagtgatttt 4860
tttttctgtt tgcaagttaa agggtttgtc attgtttctt taaaaaaaaa tacaataatg 4920
caccatatcc ctatgcataa agtgcttctt ctatttataa ggttgaaaat tctgaataac 4980
ccttttagca ttgaaaaaaa aaacaaaaac aaaaaatgga aaaaaaaaac cttgtatttt 5040
gtaaatattt tcttttcctg ctttggagct gtgtaatggc agcgaaacat gtagctgtct 5100
ttgttctata gaaatgcttt tcttcagaga agctgatctt tgttaatgtc ttgattctgt 5160
tcgcaaagca cagactagtg cttaaaaaaa aaaaagaagg aaaaattgaa aaaaatat 5218

<210> 34
<211> 763
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature

<223> Incyte ID No: 2806157CB1

<400> 34
ggcaaaacag tgacgcagca gtgtgttacc tgccgacagc ataatgtgag gcaaggtcta 60
gctgttcccc cccggcatac aagcttatag agcagcctcc tttgaagatc tccaggtgga 120
cttcacagag atgccaaagt gtggaggtgt tcgagtgtgg atcaaggact ggaacgtagc 180
ctctttgtgc ccatggtgga aaggacccca gactgtcgtc ctgatcactc ccactgctgt 240
gaacgtagag agaatcctag cctggatcca tcacaaccgt gtaaaacctg cagcgcctga 300
atcctgggag gcaagaccaa gtctggacaa cccctgcaga gtgaccctga agaagatgac 360
aagccctgct ccagtcacac ccagaagctg actggtccac gcacagccga agcatgagga 420
agctcattgt gggcttcatt tttcttaaat tttggactta cagtaagggc ttcaactgtt 480
cttactcaaa ctggggacta ttcccagtgt attcatcagg tcagtgaggt aggacagcaa 540
atgaaaacaa tctttctgtt ctatagttat tatcaatgta tgggaacgtt aaaagagact 600
tgtttgtata atgccactca gtacaaggta tgtagcccag gaaatgactg acctgatgtg 660
tgttataacc catctgagcc ccctacaacc accagttttg aaataagatt aagaactggc 720
cttttcctag gtgatacaag tgaaataata actagaacag aag 763

<210> 35
<211> 869
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5883626CB1

<400> 35
gcgagaaggg gagtggaagg caggggctga agacacaggc caggcggaat gaagatgatg 60
gtggtcttgc tcatgctgtc ctcgctcagc cggctcctgg gcctcatgag gccatcatct 120
ctcaggcaat acctggactc tgtgcccttg ccaccctgcc aggagcaaca gccaaaggct 180
agtgccgagc tagaccacaa ggcctgctac ctgtgccaca gcttgctgat gctggccggg 240
gtagttgtta gctgccagga catcactcca gaccagtggg gcgagctgca gctgctgtgc 300
atgcagttgg accgccacat cagcacgcag atccgggaga gcccccaggc catgcaccgc 360
accatgctca aggacctggc tacccagacc tacatccgtt ggcaggagct gctgacccac 420
tgccagcccc aggcccagta tttcagcccc tggaaagaca tctaaaggga cagggtcagg 480
gcagcccagg gctcctggct tcagcaggaa gtgaacaggc tcagggaact ggaggaagcg 540
aagcatcaag gccagaggag gccacatgct gaccagcctg atgaggcaag agcctgcccc 600
tgccaccgcc ccgacccctc tcctctctgc aagagcctgc ctctgccacc gccccgaccc 660
cctctcctct cagcaaggga tgggcctctc tgcctcgccc acccctcagc cctcctccca 720
gccatctcct cttccctaag gcctctgtct ccatagctct ggtttccctg ggcctcagtc 780
ctccccaccc tccttcctct gtctccctgt cactaatgtg aggtttcttt gtgcacatta 840
aagtcttctt tcagcaaaaa aaaaaaaaa 869

<210> 36
<211> 2875
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2674016CB1

<400> 36
ggggcgccat cttgtcttgt tcccgaagaa gtagaagcat cgaaagcgtt ggagaggtgt 60
taccggaacg gcggcgacaa gggtgttccc gaactagagt ggggcataca taatcttgct 120
gctatgcttc gaagctgtag tctgaatcaa cctaagtttt aaacagaagg tgaacctctg 180

agatagaaaa tcaagtatat tttaaaagaa gggatgtggg atcaaggagg acagccttgg 240
cagcagtggc ccttgaacca gcaacaatgg atgcagtcat tccagcacca acaggatcca 300
agccagattg attgggctgc attggcccaa gcttggattg cccaaagaga agcttcagga 360
cagcaaagca tggtagaaca accaccagga atgatgccaa atggacaaga tatgtctaca 420
atggaatctg gtccaaacaa tcatgggaat ttccaagggg attcaaactt caacagaatg 480
tggcaaccag aatggggaat gcatcagcaa cccccacacc cccctccaga tcagccatgg 540
atgccaccaa caccaggccc aatggacatt gttcctcctt ctgaagacag caacagtcag 600
gacagtgggg aatttgcccc tgacaacagg catatattta accagaacaa tcacaacttt 660
ggtggaccac ccgataattt tgcagtgggg ccagtgaacc agtttgacta tcagcatggg 720
gctgcttttg gtccaccgca aggtggattt catcctcctt attggcaacc aggacctcca 780
ggacctccag cacctcccca gaatcgaaga gaaaggccat catcattcag ggatcgtcag 840
cgttcaccta ttgcacttcc tgtgaagcag gagcctccac aaattgacgc agtaaaacgc 900
aggactcttc ccgcttggat tcgcgaaggt cttgaaaaaa tggaacgtga aaagcagaag 960
aaattggaga aagaaagaat ggaacaacaa cgttcacaat tgtccaaaaa agaaaaaaag 1020
gccacagaag atgctgaagg aggggatggc cctcgtttac ctcagagaag taaatttgat 1080
agtgatgagg aagaagaaga cactgaaaat gttgaggctg caagtagtgg gaaagtcacc 1140
agaagtccat ccccagttcc tcaagaagag cacagtgacc ctgagatgac tgaagaggag 1200
aaagagtatc aaatgatgtt gctgacaaaa atgcttctaa cagaaattct gctggatgtc 1260
acagatgaag aaatttatta cgtagccaaa gatgcacacc gcaaagcaac gaaagctcct 1320
gcaaaacagc tggcacagtc cagtgcactg gcttccctca ctggactcgg tggactgggt 1380
ggttatggat caggagacag tgaagatgag aggagtgaca gaggatctga gtcatctgac 1440
actgatgatg aagaattacg gcatcgaatc cggcaaaaac aggaagcttt ttggagaaaa 1500
gaaaaagaac agcagctatt acatgataaa cagatggaag aagaaaagca gcaaacagaa 1560
agggttacaa aagagatgaa tgaatttatc cataaagagc aaaatagttt atcactacta 1620
gaagcaagag aagcagacgg tgatgtggtt aatgaaaaga agagaactcc aaatgaaacc 1680
acatcagttt tagaaccaaa aaaagagcat aaagaaaaag aaaaacaagg aaggagtagg 1740
tcgggaagtt ctagtagtgg tagttccagt agcaatagca gaactagtag tactagtagt 1800
actgtctcta gctcttcata cagttctagc tcaggtagta gtcgtacttc ttctcggtct 1860
tcttctccta aaaggaaaaa gagacacagt aggagtagat ctccaacaat caaagctaga 1920
cgtagcagga gtagaagcta ttctcgcaga attaaaatag agagcaatag ggctagggta 1980
aagattagag atagaaggag atctaataga aatagcattg aaagagaaag acgacgaaat 2040
cggagtcctt cccgagagag acgtagaagt agaagtcgct caagggatag acgaaccaat 2100
cgtgccagtc gcagtaggag tcgagatagg cgtaaaattg atgatcaacg tggaaatctt 2160
agtgggaaca gtcataagca taaaggtgag gctaaagaac aagagaggaa aaaggagagg 2220
agtcgaagta tagataaaga taggaaaaag aaagacaaag aaagggaacg tgaacaggat 2280
aaaagaaaag agaaacaaaa aagggaagaa aaagatttta agttcagtag tcaggatgat 2340
agattaaaaa ggaaacgaga aagtgaaaga acattttcta ggagtggttc tatatctgtt 2400
aaaatcataa gacatgattc tagacaggat agtaagaaaa gtactaccaa agatagtaaa 2460
aaacattcag gctctgattc tagtggaagg agcagttctg agtctccagg aagtagcaaa 2520
gaaaagaagg ctaagaagcc taaacatagt cgatcgcgat ccgtggagaa atctccaagg 2580
tctggtaaga aggcaagccg caaacacaag tctaagtccc gatccaggta gtatactttt 2640
taaagtattt tgtctgattt ttaaaaaaaa ttgactgaat ttattccaag ttgaaagtgt 2700
cctttctctc tctctttaat aaactcagtt tggtacttga taaataatca tagtcttaaa 2760
tgttagaaat cctatataat attatttatt taaaattgca gatttttaat ttaaaataca 2820
tttttatttt taaattttgt cttttccctt tttttttcag atcacaaccc ctccc 2875

<210> 37
<211> 1839
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5994159CB1

<400> 37

ctcgaggggc cctctccctg ctggcacctg ggagccatgc atgaatcaag gagtcgctgg 60
acagagcctg ggtgttccca gtgctggtgc gaggacggga aggtgacctg tgaaaaggtg 120
aggtgtgaag ctgcttgttc ccacccaatt ccctccagag atggtgggtg ctgcccatcg 180
tgcacaggct gttttcacag tggtgtcgtc cgagctgaag gggatgtgtt ttcacctccc 240
aatgagaact gcaccgtctg tgtctgtctg gctggaaacg tgtcctgcat ctctcctgag 300
tgtccttctg gcccctgtca gacccccccca cagacggatt gctgtacttg tgttccaggc 360
agatggctcg gtgagctgca agaggacaga ctgtgtggac tcctgccctc acccgatccg 420
gatccctgga cagtgctgcc cagactgttc agcaggctgc acctacacag gcagaatctt 480
ctataacaac gagaccttcc cgtctgtgct ggacccatgt ctgagctgca tctgcctgct 540
gggctcagtg gcctgttccc ccgtggactg ccccatcacc tgtacctacc ctttccaccc 600
tgacggggag tgctgccccg tgtgccgaga ctgcaactac gagggaagga aggtggcgaa 660
tggccaggtg ttcaccttgg atgatgaacc ctgcacccgg tgcacgtgcc agctgggaga 720
ggtgagctgt gagaaggttc cctgccagcg ggcctgtgcc gaccctgccc tgcttcctgg 780
ggactgctgc tcttcctgtc cagattccct gtctcctctg gaagaaaagc aggggctctc 840
ccctcacgga aatgtggcat tcagcaaagc tggtcggagc ctgcatggag acactgaggc 900
ccctgtcaac tgtagctcct gtcctgggcc cccgacagca tcaccctcga ggccggtgct 960
tcatctcctc cagctccttt taagaacgaa cttgatgaaa acacagactt tacctacaag 1020
cccggcagga gctcatggtc cacactcact cgctttgggg ctgacagcca ctttcccagg 1080
ggagcctggg gcctcccctc gactctcacc agggccttcg acccctccag gagcccccac 1140
tctacctcta gcttccccag gggctcctca gccacctcct gtgactccag agcgctcgtt 1200
ctcagcctct ggggcccaga tagtgtccag gtggcctcct ctgcctggca ccctcctgac 1260
ggaagcttca gcactttcca tgatggaccc cagcccctcg aagaccccca tcaccctcct 1320
cgggcctcgc gtgctttctc ccaccacctc tagactctcc acagcccttg cagccaccac 1380
ccaccctggc ccccagcagc ccccagtggg ggcttctcgg ggggaagagt ccaccatgta 1440
aggaggtcac tgtgtccggg agactctgga gagaggacct ctgccagtgg cccagggtgt 1500
gtgcagggca gctccaagga tgaacctggt ggggatgcct gggctccctc ctgcaggggc 1560
cctggtgagg atggaagacc cccaaggctg gatgtaacct tgttcccaag aagtgtttgg 1620
aatgtgctgt aagaatggag gaagtcgttt ccactgtcag catcctccct ggaccgcgtg 1680
gctggctcat cttttgagaa gggttgggac tgccaagttc tcctggagga agagttgcgt 1740
ccggctggga ttccactcac tgggactgta ccgccaggtg tcatgcgtct ctctgaggtt 1800
tcctgattaa aggttgtctc ggtttcaaaa aaaaaaaaa 1839

<210> 38
<211> 1232
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2457335CB1

<400> 38
gggcagcctg cgcctgggta ccgaggctgc tgcgcggcgg acagcgggcg cgatgtatct 60
ccgcagggcg gtctccaaga ctctggcgct gccgctgagg gcgccccccca accccgcgcc 120
gctcggaaag gacgcatctc tgcgccggat gtcatctaac agattccctg gatcatctgg 180
atcaaatatg atttattatc tggttgtagg cgtcacagtc agtgctggtg gatattatgc 240
ttacaagaca gtaacatcag accaagccaa acacacagaa cataaaacaa atttgaaaga 300
aaaaacaaaa gcagagatac atccatttca aggtgaaaag gagaatgttg cggaaactga 360
gaaagcaagt tcagaagccc cagaagaact tatagtggaa gctgaggtgg tagatgctga 420
agaaagtccc agtgctacag ttgtggtcat aaaagaggca tctgcctgtc caggtcacgt 480
ggaggctgct ccggagacca cagcagtcag tgctgaaacc gggccagagg tcacagatgc 540
agcggcgagg gaaaccacgg aagtaaaccc tgaaacgacc ccagaggtta caaatgctgc 600
cctggatgaa gctgtcacca tcgataatga taaagataca acaaagaacg aaacctctga 660
tgaatatgct gaactagaag aagaaaattc tccagctgag tcagagtcct ctgctggaga 720
tgatttacag gaggaagcca gtgttggctc tgaggctgct tcggctcaag gcaatctcca 780
gccggtagac atttcagcaa caaatgccat agggtgtctg ataagtgctt tggtgttttt 840

agtacactta gtttaaaaaa aaaaagactt attttctaga aaacgttaag gggtctggaa 900
gattttgggc atccactgat agattttagg actgagagac ttgagattgg tacatttctt 960
actactcttc tacggtctga gatcagaata tgacattgca gtaaagagct taaatagttt 1020
catcttgggt tttttttct agacactttt cttttatccc agtcatctct gaatttacat 1080
tttatattaa agaaatggag ctatccataa aacctgtatt tactttaatt tcaccatgct 1140
ttcagcttac tttatggtat gatatgtgag gctaacatat agttggggtc caataaaatt 1200
tgaaaacgtg taaaaaaaaa aaaggggggc cg 1232

<210> 39
<211> 3250
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2267802CB1

<400> 39
ggcccccgcc caggtgtctc cctttgggaa gctgcccgcc gagtctccga gatttgtccc 60
tggtggtccc gcggacccct cgtccctccg cagtctccgg ctggcagcga tggagggcgc 120
tggggagaac gccccggagt ccagctcctc tgcccctggg tccgaagagt ctgccaggga 180
tccacaggtg ccgcctccgg aggaagaatc gggggactgc gcccggtccc tggaggcggt 240
ccccaagaaa ctctgtgggt atttaagtaa gttcggcggc aaagggccca tccggggctg 300
gaaatcccgc tggttcttct acgacgaaag gaaatgtcag ctgtattact cgcggaccgc 360
tcaggatgcc aatcccttgg acagcatcga cctctccagt gcagtgtttg actgtaaggc 420
ggacgctgag gaggggatct tcgaaatcaa gactcccagc cgggttatta ccctgaaggc 480
cgccaccaag caagcgatgc tgtactggct gcagcagctg cagatgaagc gctgggaatt 540
ccacaacagc ccgccggcac ctcctgccac ccctgatgcc gccctggctg ggaatgggcc 600
cgtcctgcac ctcgagctag ggcaagaaga ggcagagctg gaggagttcc tgtgccctgt 660
gaaaacaccc cctgggctag tgggcgtggc agctgccttg cagcccttcc ctgcccttca 720
gaatatttcc ctcaagcacc tggggactga aatacagaac acaatgcaca acatccgtgg 780
caacaagcag gcccagggaa caggccatga acctccaggg gaagattcta cacagagtgg 840
ggagcctcag agggaggagc agccctcggc ctctgacgcc agcaccccag tgagagagcc 900
agaggattct ccaaagcctg cacccaagcc ttctctgacc atcagtttcg ctcagaaagc 960
caagcgccag aacaacacct tcccattctt ttctgaagga atcacacgga accgaactgc 1020
ccaggagaaa gtggcagcct tggagcaaca ggttctgatg ctcaccaagg agttaaagtc 1080
tcagaaggag ctagtgaaga tcctgcacaa ggcactggag gccgcccagc aggagaagcg 1140
ggcgtccagc gcatacctgg cggcggctga ggacaaggac cggctggagc tggtgcggca 1200
caaagtgcgg cagatcgcgg agctgggccg gcgggtggag gccctggagc aggagcggga 1260
gagcctggcg cacacagcga gcctgcggga gcagcaggtg caggagctac agcagcacgt 1320
gcagctgctt atggacaaga accacgccga gcagcaggtc atctgcaagc tctctgagaa 1380
ggtcacccag gacttcacgc acccccctga ccagtctcct ttgcgccccg acgctgccaa 1440
cagggacttc ctgagccagc aggggaagat agagcacctg aaggatgaca tggaagctta 1500
ccggacccag aactgcttcc tcaactccga gatccaccag gtcacaaaga tctggagaaa 1560
ggtggctgag aaggagaagg cccttctgac gaagtgcgcc tacctccaag ccagaaactg 1620
ccaggtggaa agcaagtacc tggccggtct gagaaggctg caggaggccc tgggggacga 1680
agccagcgag tgctcagagc tgctgaggca gcttgtccag gaggcactgc agtgggaagc 1740
tggggaggcc tcatctgaca gcatcgagct gagccccatc agtaagtatg atgagtacgg 1800
cttcctgacg gtgcccgact atgaggtgga agacctgaag ctgctggcca agatccaggc 1860
attggagtca cgatcccacc acctgctggg cctcgaggct gtggatcggc cgctgaggga 1920
gcgctgggct gccctgggcg atcttgtgcc ctcagccgag ctcaagcagc tactgcgggc 1980
aggagtaccc cgtgaacacc ggcctcgtgt ctggaggtgg ctggtccacc tccgtgtcca 2040
gcacctgcac actccaggct gctaccagga actgctgagc cggggccagg cccgcgagca 2100
ccctgctgcc cgccagattg agctggacct gaaccggacc ttccccaaca acaaacactt 2160
cacctgcccc acctccagct tccccgacaa gctccgccgg gtgctgctgg ccttctcctg 2220
gcagaacccc accatcggct actgccaggg cctgaacagg ctggcggcca ttgccctgct 2280

ggtcctagag gaggaggaga gcgccttctg gtgcctggtg gccattgtgg agaccatcat 2340
gcccgctgat tactactgca acacgctgac ggcatcccag gtggaccagc gggtgctcca 2400
ggacctgctc tcggagaagc tgcccaggct gatggcccat ctggggcagc accacgtgga 2460
tctctccctc gtcaccttca actggttcct cgtggtcttt gcggacagtc tcattagcaa 2520
catcctcctt cgggtctggg atgccttcct gtacgagggg acgaaggtgg tgtttcgcta 2580
tgccttggcc attttcaagt acaacgagaa ggagatcttg aggctacaga atggcctgga 2640
aatctaccag tacctgcgct tcttcaccaa gaccatctcc aacagccgga agctgatgaa 2700
catcgccttc aatgacatga acccettccg catgaaacag ctgcggcagc tgcgcatggt 2760
ccaccgggag cggctggagg ctgagctgcg ggagctggag cagcttaagg cagagtacct 2820
ggagaggcgg gcatcccggc gcagagctgt gtccgagggc tgtgccagcg aggacgaggt 2880
ggaggggggaa gcctgacttg gccacctccc ctccccacag ccttcctcac ccttggctgg 2940
cagacccact ggaggtcagg cacggaccag tggcccagcc ctgggtgtcc catcaccatg 3000
tgaccttgga catgtccctt cccctctctg gccctcagtt tccccactgg gacattgtgt 3060
gctgcaaagc cattggttgg gctacttctt cataggcact tacttaccca gggatgccac 3120
cctttcgtca cctcttccac agagcacttt ggcatgtaaa caagcaagag cactgcctct 3180
atagggtaac ctggaacatt ctctaggtta tatcaatata aaacaatgta aatggtggaa 3240
aaaaaaaaaa 3250

<210> 40
<211> 3621
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3212060CB1

<400> 40
aaggcgagta gcatgtgcgg gagactcacg ttgccggcga agtgggagag agaaaagtgg 60
taacctgggg ctggggggccg gcgcggcgga gctcggagta gtagagcgga gtgaagacac 120
gggggaggat agagactggc attcctttgg gccgggggat tggcgggagt cgtgctgggt 180
gctctcgccg tgttgaggtc ccagtgaggg gaaggagaag cggaagaggg tctctagtcg 240
gggcctaggg caaagggact acaaaaagga tgcagatgac tatagaaatg aggacgacga 300
ggagatgctg tggaggagca gtagaggtga gaagatgatg caaagaaact gtgtcagtga 360
ggaactgtat agagggtcat agaggtgagg tggcggagag aaactaacta acggaccata 420
gaggtggggg agccattgta gaaggacgtg gacgcgaaag ggtcgtgtag atgggcatat 480
gtgtgaagca gcaacgtaga ggggctgaag aggagaaatt catggagaga aagaatgcac 540
ctagagtgag ctctgcagag tgctgcgtgg gatatcccta gagtttggtc tagtgaaggc 600
acgctaacca ggcacctaag gcatttcaag tagtgacttc ccacatttgg ctaggaatgt 660
gggtcctcct ccgaagtggg taccccctcc gtatcttgtt acccctgcgt ggggagtgga 720
tgggtcggag gggcctgccc cgaaacttgg ccccaggccc tcctcgcaga cgttacagga 780
aggagactct ccaagccttg gatatgccag tgttgcctgt aactgcaact gaaatccgcc 840
agtatttgcg gggcatggg atccccttcc aggatggtca cagttgcctg cgggcactga 900
gcccctttgc agagtcttca cagctcaaag gccagactgg tgttaccact tccttcagcc 960
tcttcattga caagaccaca ggccactttc tctgcatgac cagcctagca gaagggagct 1020
gggaagactt ccaggccagc gtggaggggc gaggggatgg ggccagggag gggtttctgc 1080
ttagcaaggc accagaattt gaggacagcg aggaggtccg gaggatctgg aaccgagcaa 1140
tacctctctg ggagctgcct gatcaggagg aggttcagct ggctgataca atgtttggcc 1200
ttaccaaggt tacagatgac acactcaagc gtttcagtgt gcgatatctg cgacctgctc 1260
gcagtcttgt cttcccttgg ttctcccctg gggggttcagg attacgaggc ctgaagctcc 1320
tagaggctaa atgccagggg gatggagtga gctacgagga aaccactatt ccccgaccca 1380
gcgcctacca caatctgttt ggattaccac tgattagtcg tcgagatgct gaggtggtac 1440
tgacgagtcg tgagcttgac agcctggcct tgaaccagtc cacggggctg cctacccctta 1500
ctctaccccg aggaacgacc tgcttacccc ctgccttact cccttacctg gaacagttcc 1560
ggcggattgt attctggttg gggatgacc ttcggtcctg ggaagccgcc aagttgtttg 1620
cacgaaaact gaaccccaaa cgatgcttct tggtgcgacc aggagaccag caaccccgtc 1680

ccctggaggc cctgaacgga ggcttcaatc tttctcgtat tcttcgtacc gccctgcctg 1740
cctggcacaa gtccatcgta tctttccggc agcttcggga ggaggtgcta ggagaactgt 1800
cgaatgtgga gcaagcagct ggcctccgct ggagccgctt tccagacctc aatcgtatct 1860
tgaagggaca tcgaaagggc gagctgacgg tcttcacagg gccaacaggc agtggaaaga 1920
cgacattcat cagtgagtat gccctggatt tgtgttccca gggggtgaac acactgtggg 1980
gtagcttcga gatcagcaat gtgagactag cccgggtcat gctgacacag tttgccgagg 2040
ggcggctgga agatcaactg gacaaatatg atcactgggc tgaccgcttt gaggacctgc 2100
ccctctattt catgactttc catggacagc aaagcatcag gactgtaata gatacaatgc 2160
aacatgcagt ctacgtctat gacatttgtc atgtgatcat cgacaacctg cagttcatga 2220
tgggtcacga gcagctgtcc acagacagga tcgcagctca agactacatc atcggggtct 2280
ttcggaagtt tgcaacagac aataactgcc atgtgacact ggtcattcac ccccggaaag 2340
aggatgatga caaggaactg cagacagcgt ccatttttgg ctcagccaaa gcaagccagg 2400
aagcagacaa tgttctgatc ctgcaggaca ggaagctggt aaccgggcca gggaaacggt 2460
atctgcaggt gtccaagaac cgctttgatg gagatgtagg tgtcttcccg cttgagttca 2520
acaagaactc cctcaccttc tccattccac caaagaacaa ggcccggctc aagaagatca 2580
aggatgacac tggaccagtg gccaaaaagc cctcttctgg caaaaagggg gctacgacac 2640
agaactctga gatttgctca ggccaggccc ccactcccga ccagccagac acctccaagc 2700
gttcaaagtg aaggccgtgc agagctggtc actgaaatga gcctgatagg ataggctgga 2760
gcataaaact ctgcaagggc tcctctatcc tgtggtcctg agctgtgtgc ccttctcagt 2820
ctgaggggcc taacctagag caggtttcca tagtgagaaa attcaatgta gcagactact 2880
gagaaactac tgtgttgctc aggctttgtt tgaggtcctg tatatacagc actgaaaaga 2940
gagataaagt ccctgcctgc atgcattctg gcggaagaga caagcaagca atgaacaaat 3000
tagcagaaaa cctagtttta gtgaaaaatg ctgtaaagaa aatagaaatg cgatagagtg 3060
ctggcaggct agtgtagata agtggtctga aaggtgtct ctgagccgag ggcatgtgag 3120
ctggggccta aacaactaga aggagagagc cacgtgaaca tccggcgaag gggacccagg 3180
cagagagaaa aggaaatcca agccctgagg taggaatgag caggtcagat tcaaggcagt 3240
gaggtcaggc cgcatgaacc tggaggggaa tcggggactt catgcgaaac tccagcctag 3300
gctttcaaag tcaaagggtg atacagtggg taccaagctc ctctgctccc cactttgtag 3360
agcctagcat gaggtggcat gtactagaat tggatcctag gtgcttagcc ctgcaatatc 3420
agggcctcac tggtgggagc tgcctcgggc tgggttgctt ggtcatagag ccatagaagg 3480
aagctgtcag cccggagtgc ctgccaccta gacactgatg ccattgtgtg ctgcctcaag 3540
actgctggag tcaggacatt ttatagagcc ttttccagtt ttactaaaaa atttttccat 3600
tgaaaaaaaa aaaaaaaaaa a 3621

<210> 41
<211> 1693
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3121069CB1

<400> 41
catgaatgta ctgcaaggga acacatttgt gtcatgtgaa gagacatgac aaaaacagcc 60
ctccttaaat tatttgtggc aatagtgatc acattcattt taattttgcc ggaatatttc 120
aagacaccga aagaaagaac attggagcta tcatgtctgg aagtgtgttt gcaatctaat 180
tttacctatt cactctcctc cttaaatttt tcttttgtga cttttctgca accagtaagg 240
gaaactcaga ttatcatgag aatctttcta aatccctcca attttcgtaa cttcaccagg 300
acttgccaag acatcacagt tcttatcagg agaggatcaa tggaagtgaa agcaaatgat 360
tttcattcac cttgtcagca ctttaacttc agtgtagctc ctctggttga ccacttggag 420
gaatataaca ctacctgtca tctaaaaaac cacactggaa gatcaacaat catggaggat 480
gagccaagca aggagaaatc gataaactac acttgtagaa tcatggaata cccgaatgat 540
tgtatacaca tttctttgca cttggagatg gatataaaaa atatcacttg ttccatgaag 600
atcacttggt atattttagt tctattagtt tttatatttt tgatcatcct cactatccgc 660
aaaatacttg aaggccagag aagagtgcaa aagtggcaga gtcatagaga caaacctaca 720

tctgttctct taagaggaag tgattcggag aaactgagag cattgaatgt gcaggttctt 780
tcagagacca cgcagaggct gcctttggat caagtccagg aagtgcttcc cccaattcca 840
gaactataag ttacttccac agtgcatcag tgagatcaat atacacgaat atccccgggc 900
aagttggacc gagccctttg aagaatactc agaagtttat tttgtgaatg agtagactgg 960
aaaatgtttg tgtccagctg aggatgcaca gttggaaagc aggaggaatg ctgactggtt 1020
gatgaaaact agcttaagag cattcattcg ctccatgaga tcaagggaac aagagtgttt 1080
gcaagaagcc attatgagtc atggaaaaaa agatgatgaa acccatggaa acagcaagag 1140
aattccccact ctctctcttc ttaaaaaaaa tctatcatta tacagcacag agtggagcca 1200
agtttttaat tttgaggaac caaaaacagg atcaaatatg aaaacccttt cttttattgg 1260
gccacattgt agatgctgat ttgataattg tttcctatgc agatagatta tttttatttc 1320
acagattatt taaaagggaa gagggcctgg ttgtttattt atatgtttgt ttgcatttat 1380
gaatcttgct gcctttttagc accaggatgt ttttaaaaaa attcaaagag gccaggcgca 1440
gtggctcatg cctgtaatcc cagcactttg ggattctgag gtgggaggat catgaggtca 1500
agggatcgag accatcctgg ccaacatggt gaaaccctgt ctctactaaa aacacaaaaa 1560
ttagctgggt gtggtggtgc gcgcctgtag tcccagctcc tcaggaggct gaggcaggag 1620
aatcacttga acctggcagg cagagtttgc agtgaaccaa gatcacgcca ctgcattaca 1680
gcctagacaa gca 1693

<210> 42
<211> 2289
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3280626CB1

<400> 42
ggccgctgta acctcttcgg tccgcgacga tcctctagag cactgtgtgt ctccccggac 60
gcgagcccgc tcccctgagt aagagtcagc cagccgcgga tggggagcgt gagtggcgag 120
aatctgcaaa atggctgata atttggatga atttattgaa gagcaaaaag ccagattggc 180
cgaagacaaa gcagagttgg aaagtgatcc accttacatg gaaatgaagg gaaagttgtc 240
agcgaagctt tctgaaaaca gtaagatact gatctctatg gctaaggaaa acataccacc 300
aaatagtcaa cagaccaggg gttccttagg aattgattat ggattaagtt taccacttgg 360
agaagactat gaacggaaga aacataaatt aaaagaagaa ttgcggcaag attacagacg 420
ttatcttact caggaaaggt tgaaacttga acgtaacaaa gaatacaatc agtttctcag 480
gggtaaggaa gaatccagtg aaaagttcag gcaggtggaa aagagtactg agcccaagag 540
tcagagaaat aaaaaaccta ttggtcaagt taagcctgat ctaacttcac aaatacagac 600
atcttgtgaa aattcagagg gtcctagaaa agatgtctta actccttcag aggcatatga 660
agaacttctg aaccaaagac gactagagga ggacagatac cgacaactag atgatgaaat 720
cgaattaagg aatagaagaa ttattaaaaa agcaaatgaa gaagtgggca tttccaacct 780
aaaacatcaa aggtttgcaa gcaaggctgg cattccagat agaagatttc acagatttaa 840
tgaggatcgt gtttttgata gacggtatca tagaccagac caagatcctg aagtaagtga 900
agaaatggat gagaggttta gatatgaaag tgattttgat agaagacttt cgagagtgta 960
tacaaatgac aggatgcaca ggaacaaacg agggaatatg cctcctatgg aacatgatgg 1020
ggatgttata gaacagtcaa acataagaat ttcatctgct gaaaataaaa gtgctccaga 1080
caatgaaaca tccaaatctg ctaatcaaga tacctgtagt ccttttgcag ggatgctctt 1140
tggaggtgaa gatcgagaac ttattcagag aaggaaagag aaatacagac tagaactgtt 1200
ggaacaaatg gctgagcaac agaggaacaa gagacgagaa aaagatttag aactcagggt 1260
tgcagcgtct ggagcacaag accctgagaa atcgcctgat agactaaagc agtttagtgt 1320
ggcaccaaga cactttgaag agatgatacc acctgaaaga cccagaatag ctttccagac 1380
acctctccct cctttatctg ccccatctgt cccacccatc ccatcagttc atcctgttcc 1440
ttctcaaaat gaagatttgc gcagtggact cagcagcgcc cttggtgaaa tggtgtctcc 1500
caggattgca cctctgcctc cacctcccct actaccacct ttggctacta actatcgaac 1560
tccttatgat gatgcatact atttttatgg gtccaggaat actttcgatc ccagtcttgc 1620
ttattatggt tcaggaatga tgggcgtaca gcctgcagct tatgttagtg ctcctgtcac 1680

ccaccaacta gcacaacctg ttgtagtctc accctgtcac ccaggctgga gtacaatgtt 1740
gtgaacttgg ctcactgcaa cctccgcctc ctgagttcaa gtgattcttc tggctcagcc 1800
tcctgagaaa ctgggattaa aggcgtgcac cactatgccc ggctaagttt ttgtatcttt 1860
agtagagaca gggtttcacc atgttggcca ggctggtctc gaactcctga ccttgtgatc 1920
tgcctgcctc agcctcccaa agtcctggga ttacaggaat acggttggac agaatgaact 1980
gaagattaca agtgatcaag tgataaattc aggattgatt tttgaagata aaccgaaacc 2040
ttccaaacag tcacttcagt cttaccaaga ggctttgcag cagcagattc gggaaaaaaa 2100
aaaaaagagg ggggcggcga atattgagct cgtgaccgcg gaataaattc gggcgcgaac 2160
ctgcaggcga gagggaggga atctatatca agatatcgat accgggaccc gaaggggggc 2220
gcggacccaa ttcgctaagg gagcgataag gcggcaaggc gcggtaaacg cgggtggaac 2280
ctgcgtcac 2289

<210> 43
<211> 1304
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 484404CB1

<400> 43
ctcgcccacg cgtccggggc aggtaacagc tgcatcattg accgcacagc gccatctctc 60
cctgagaata aagccgatag ccaccctcct ccggctccga gcctgcttct gccacacctc 120
gctctcagtt ctctccacat ttccatagag accgtgtggt ttttgttcac ccgggccccc 180
catttatagg cataaaatcc actgtctgcc agcctcccttt ccctcccacc tttttgtttt 240
acattttttta caccaatgta ccaaaaaggc ggacggctgc atttacgggg tctcccggag 300
ggccagagtc gtggcttaca gaagagacga aatgtggtct gagggacgat atgaatatga 360
aagaattccg agagaacgag cacctcctcg aagtcatccc agtgatgaat ctggttatag 420
atggacaaga gacgatcatt ctgcaagcag gcaacctgaa tacagggaca tgagagatgg 480
ctttagaaga aaaagtttct actcttccca ttatgcgaga gagcggtctc cttataaaag 540
ggacaatact tttttcagag aatcacctgt tggccgaaag gattctccac acagcagatc 600
tggttccagt gtcagtagca gaagctactc tccagaaagg agcaaatcat actctttcca 660
tcagtctcaa catagaaagt ccgtgcgtcc tggtgcctcc tacaaacggc agaatgaagg 720
aaatcctgaa agagataaag agaggcctgt ccagtctttg aaaacatcaa gagatacttc 780
accctcaagt ggttcagcag tttcttcatc aaaggtgtta gacaaaccca gtaggctaac 840
tgaaaaggaa cttgctgagg ctgcaagcaa gtgggctgct gaaaagctag agaaatcaga 900
tgaaagtaac ttgcctgaaa tttctgagta tgaggcggga tccacagcac cattgtttac 960
tgaccagcca gaggaacctg agtcaaacac aacacatggg atagaattat ttgaagatag 1020
tcagctaacc actcgctcta aagcaatagc atcaaaaacc aaagagattg aacaggttta 1080
ccgacaagac tgtgaaactt tcgggatggt ggtgaaaatg ctgattgaaa aagatccttc 1140
attagaaaag tctatacagt ttgcattgag gcagaattta catgaaatag gtgagcggtg 1200
tgttgaagaa ctcaagcatt tcattgcaga gtatgatact tccactcaag attttggaga 1260
gcctttttag atttttctgc tcaggctaaa aaaaaaaaaa aagg 1304

<210> 44
<211> 4850
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2830063CB1

<400> 44
gtgcagcggc tgagatcacg tggtgcgccg ggaagccacc ccgcctctcc gaggcctccc 60

tgccccgccc cgtcacgccc ctttcccggc gggacgcttt gagccgcccc gaactaagca 120
gggcggtcgg gggagtcata ctccatgggt ttatgtgata aatgatcttg gcatagaaga 180
gaatttaaag aatgatggct tcattccagc gctccaatag tcatgacaaa gtaaggagaa 240
tagttgcaga ggagggtcgt acagcaagaa acctaatagc ttggagtgtt ccactagaaa 300
gcaaagatga tgatggaaaa cctaaatgtc aaactggtgg aaaatctaag aggaccattc 360
aaggcactca taaaactact aaacagagta ctgcagtgga ctgtaaaata acatcgtcta 420
cgactggaga taaacacttt gataaaagtc ccactaaaac aaggcaccct cggaaaattg 480
atctaagagc tcgatactgg gcatttcttt ttgataatct tcgccgagca gtagatgaaa 540
tctatgtaac ttgcgaatca gatcagagtg tggtcgaatg taaggaggtg ctaatgatgc 600
tggataacta tgtaagagat ttcaaagcat tgattgactg gattcagctt caggaaaagc 660
tagagaagac agatgctcaa agcagaccaa catcattggc atgggaagta aagaagatgt 720
ctccgggacg ccatgtgatt ccaagtccat caacagatag aataaatgta acatcaaatg 780
ctcgacgaag cttaaatttt ggaggttcaa ctggcacagt gccagctcct cgtctggctc 840
ccacaggtgt cagttgggct gacaaggtaa aggctcatca tacaggctct actgcttctt 900
cagaaataac acccgcccag tcttgcccac caatgacagt gcagaaggcc tcacgcaaaa 960
atgaacggaa agatgctgaa ggatgggaga ccgttcagag aggaaggcct attcgttctc 1020
gatcaacagc agtgatgcca aaagtttcat tggcaacaga agccacaaga tcaaaggatg 1080
acagtgataa agaaaatgta tgtctttac ctgatgaaag catacagaaa ggtcaatttg 1140
ttggagatgg aacttctaat actatagaat ctcatcccaa agactcatta cactcttgtg 1200
accatcctct tgccgaaaaa acccagttca cagtgagtac attggatgat gtgaagaatt 1260
ctggcagtat tcgagacaat tatgttcgaa cttctgaaat atctgctgtc cacattgata 1320
cagagtgtgt ttcagttatg ctgcaagctg gtacacctcc tttacaagta aatgaagaaa 1380
aatttccagc agagaaagca aggatagaaa atgaaatgga cccttcagat atttcaaatt 1440
ccatggcaga agtccttgct aaaaaagaag agctagcaga tcgtctagaa aaggccaatg 1500
aagaagccat tgctagtgct attgctgaag aagaacagtt aactagagaa attgaagctg 1560
aagaaaacaa tgatattaac attgaaactg acaacgacag tgatttttct gccagcatgg 1620
gcagtgggag tgtttctttc tgtggtatgt ccatggactg gaacgatgtc cttgcagatt 1680
atgaagctcg tgagtcttgg cgccaaaata catcctgggg ggacattgta gaagaagaac 1740
ctgctagacc tccagggcat ggaattcaca tgcatgaaaa actttcttca ccctctcgta 1800
aaagaacaat tgcagaatct aagaagaaac atgaagaaaa acaaatgaaa gcacagcagc 1860
taagggaaaa gttacgcgaa gagaaaacat tgaagcttca gaaattgtta gaaagggaga 1920
aggatgtccg gaagtggaag gaagaattgc tagatcaacg acgcaggatg atggaagaaa 1980
aattacttca tgctgagttt aagcgagaag tgcagttaca agcaattgtg aaaaaagcac 2040
aagaagaaga agctaaggta aatgaaattg cctttataaa taccccttgaa gcccagaata 2100
aacgtcatga tgttttatca aaattgaagg aatatgaaca gaggcttaat gagctacagg 2160
aagagcgtca gagaagacag gaagaaaagc aagcacgtga tgaagctgtg caggaacgca 2220
agagagctct agaggcagag cggcaggccc gtgtagaaga attgttaatg aagaggaaag 2280
aacaagaagc ccgaattgaa caacagaggc aagaaaagga aaaagcccgt gaggatgcag 2340
cccgggaaag agctagagac agggaagaac gattggcagc actcacagct gctcaacaag 2400
aagctatgga agagttacag aaaaaaattc agctcaagca tgatgaaagt attcgaaggc 2460
acatggaaca gattgaacaa agaaaagaaa aagctgctga gctaagcagt gggcgacatg 2520
caaatactga ttatgccccc aaactgaccc cttatgaaag aaagaagcag tgttctctct 2580
gcaatgtcct gatctcttca gaggtatatc tttttagcca tgttaaaggg agaaaacacc 2640
agcaagccgt gagagagaat accagcatcc aggggcgtga actgtcagat gaagaagtgg 2700
agcatctttc cttgaagaag tacattattg acattgtggt tgaaagtaca gctccagcag 2760
aagctttgaa agatggagaa gagcggcaaa aaaataaaaa aaaagccaaa aagataaaag 2820
cccggatgaa cttcagggct aaggaatatg agagtttaat ggaaaccaaa aattctggct 2880
ctgattcacc ttataaagca aagcttcagc gattagccaa agatcttcta aaacaagtac 2940
aagttcaaga cagtggctca tgggcaaaca ataaagtgtc tgctttggat cggaccctag 3000
gagagatcac tagaatactg gaaaaagaga atgtggcaga tcagattgca tttcaagctg 3060
ctggtggatt aacagccctt gaacacatcc ttcaagcagt agtcccagcc acaaatgtga 3120
acacagtttt aagaattcct cctaagtctc tctgcaatgc aatcaatgtt tacaacctca 3180
cctgcaataa ctgttcagaa aactgcagtg atgttctgtt tagtaacaag attaccttct 3240
taatggacct cctgatacac cagttgacgg tttatgttcc agatgaaaat aatactattt 3300
tggggagaaa tacaaataaa caagtttttg aaggcttgac aactggactt ctcaaagtca 3360
gtgctgtggt tttgggctgc ctgattgcca atcgaccaga tggaaactgc cagccagcta 3420

ccccaaaaat accaacacag gaaatgaaaa acaaaacctc acaaggtgat ccttttaaca 3480
atcgagttca ggaccttatc agctacgtgg tgaacatggg tctgattgac aaactgtgtg 3540
cctgcttcct ctcggtgcaa ggcccagtgg atgagaatcc caagatggcc atatttctgc 3600
agcatgccgc aggactctta catgcaatgt gtacactgtg ctttgctgtc actggaaggt 3660
catacagcat atttgacaat aatcgccagg atcccacagg gctgacagct gctcttcagg 3720
caaccgacct ggctggagtt cttcatatgc tctactgtgt cctcttccat ggcaccatct 3780
tggaccccag cactgccagt cccaaggaga attacactca aaataccatc caagtggcca 3840
ttcagagttt acgtttcttc aacagctttg cagctcttca tctgcctgct tttcagtcta 3900
ttgtaggggc agagggcttg tcccttgcat tccggcacat ggccagctcc ctgctgggcc 3960
actgcagcca agtctcctgt gaaagcctcc ttcatgaggt catcgtctgt gtgggctact 4020
tcactgtcaa ccacccagat aaccaggtga tcgtgcagtc cggccgccac cccacagtgc 4080
tgcagaagct ctgccagttg cccttccagt atttcagtga cccacggctg atcaaagtac 4140
tgttcccttc acttatcgct gcttgttaca acaaccatca gaacaagatc attctggagc 4200
aagagatgag ctgtgtttta ctggccactt tcattcagga tttggcacag actccaggtc 4260
aagcggaaaa ccagccttac caacccaaag ggaaatgcct tggttcccaa gactatcttg 4320
agctggctaa cagatttcct cagcaggcct gggaagaagc tcgacagttt ttcttgaaaa 4380
aagagaaaaa ataaatgttt tggttgattc tgtatttgag taccccttgtt aatattttaa 4440
attgtccaaa caaacattct aattgttcct taagaactca ttttcccatg tttatactct 4500
tcccacactg tagatatggc atgtacttta cactatttat aatgactgta gatacttgaa 4560
tgttctactt gctaattttg caagttgagt ttatttcatt tatgcagagt atcttggagt 4620
tggtaatttc catcttatga taatatatac tttgcatttg tgatatgggt gaaaggagac 4680
ataaaattag caagtctgtt tgttcttgta ataaagtaac ttattctgtt ttcattgttg 4740
acttttcatg ttaaggaaat acgaatctga aagaaaaatg ttaactccag ctcttgaagt 4800
atcttaaata aagacttaat taaagtttaa caaaaaaaaa aaaaaaaaaa 4850

<210> 45
<211> 4350
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7506096CB1

<400> 45
atggaatcta gttcatcaga ctactataat aaagacaatg aagaggaaag tttgcttgca 60
aatgttgctt ccttaagaca tgaactgaag ataacagaat ggagtttgca gagtttaggg 120
gaagagttat ccagtgttag tccaagtgaa aattctgatt atgcccctaa tccttcaagg 180
tctgaaaagc taattttgga tgttcagcct agccaccctg gacttttgaa ttattcacct 240
tatgaaaacg tctgtaaaat atctggtagc agcactgatt ttcaaaaaaa gccaagagat 300
aagatgtttt catcttctgc ccctgtggat caggagatta aaagccttcg agagaaacta 360
aataaactta ggcaacagaa tgcttgtttg gtcacacaga atcattcctt aatgactaaa 420
tttgaatcta ttcactttga attaacacag tcaagagcaa aagtttctat gcttgagtct 480
gctcaacagc aggcagccag tgtcccaatc ttagaagaac agattataaa tttggaagca 540
gaggtttcag ctcaagataa agttttgaga gaggcagaaa ataagctgga acagagccag 600
aaaatggtaa ttgaaaagga acagagtttg caggagtcca aagaggaatg tataaaatta 660
aaggtggact tacttgaaca aaccaaacaa ggaaaaagag ctgaacgaca aaggaatgaa 720
gcactatata atgccgaaga gctgagtaaa gctttccaac aatataaaaa aaaagtggct 780
gaaaaactgg aaaaggttca agctgaagaa gaaatattag agagaaatct aactaactgt 840
gaaaaagaaa ataaaaggct acaagaaagg tgtggtctat ataaaagtga acttgaaatt 900
ctgaaagaga aattaaggca gttaaaagaa gaaaataaca acggaaaaga aaaattaagg 960
atcatggcag tgaaaaattc agaagtcatg gcacaactaa ctgaatctag acaaagtatt 1020
ttgaagctag agagtgagtt agagaacaaa gacgaaatac ttagagacaa attttcttta 1080
atgaatgaaa accgagaatt aaaggtccgt gttgcagcac agaatgagcg actagattta 1140
tgtcaacaag aaattgaaag ttcaagggta gaactaagaa gtttggaaaa gattatatcc 1200
cagttgccat taaaaagaga attatttggc tttaaatcat atctttctaa ataccagatg 1260

agtagcttct caaacaagga agaccgttgc attggctgct gtgaggcaaa taaattggtg 1320
atttcggaat tgagaattaa gcttgcaata aaagaggcag aaattcaaaa gcttcatgca 1380
aacctgactg caaatcagtt atctcagagt cttattactt gtaatgacag ccaagaaagt 1440
agcaaattaa gtagtttaga aacagaacct gtaaagctag gtggtcatca agtagcagaa 1500
agcgtaaaag atcaaaatca acatactatg aacaagcaat atgaaaaaga gaggcaaaga 1560
cttgttactg gaatagaaga actacgtact aagctgatac aaatagaagc tgaaaattct 1620
gatttgaagg ttaacatggc tcacagaact agtcagtttc agctgattca agaggagctg 1680
ctagagaaag cttcaaactc cagcaaactg gaaagtgaaa tgacaaagaa atgttctcaa 1740
cttttaactc ttgagaaaca gctggaagaa aagatagttg cttattcctc tattgctgca 1800
aaaaatgcag aactagaaca ggagcttatg gaaaagaatg aaaagataag gagtctagaa 1860
accaatatta atacagagca tgagaaaatt tgtttagcct ttgaaaaagc aaagaaaatt 1920
cacttggaac agcataaaga aatggaaaag cagattgaaa gagttaggca actagattca 1980
gcattggaaa tttgtaagga agaacttgtc ttgcatttga atcaattgga aggaaataag 2040
gaaaagtttg aaaaacagtt aaagaagaaa tctgaagagg tatattgttt acagaaagag 2100
ctaaagataa aaaatcacag tcttcaagag acttctgagc aaaacgttat tctacagcat 2160
actcttcagc aacagcagca aatgttacaa caagagacaa ttagaaatgg agagctagaa 2220
gatactcaaa ctaaacttga aaaacaggtg tcaaaactgg aacaagaact tcaaaaacaa 2280
agggaaagtt cagctgaaaa gttgagaaaa atggaggaga aatgtgaatc agctgcacat 2340
gaagcagatt tgaaaaggca aaaagtgatt gagcttactg gcactgccag gcaagtaaag 2400
attgagatgg atcagtacaa agaagagctg tctaaaatgg aaaaggaaat aatgcaccta 2460
aaacgagatg gagaaaataa agcaatgcac ctctctcaat tagatatgat cttagatcag 2520
acaaagacag agctagaaaa gaaaacaaat gctgtaaagg agttagaaaa gttacagcac 2580
agtactgaaa ctgaactaac agaagccttg caaaaacggg aagtacttga gactgaacta 2640
caaaatgctc atggagaatt aaaaagtact ttaagacaac tccaggaatt gagagatgta 2700
ctacagaagg ctcaattatc attagaggaa aatacacta ctataaagga tctcacagct 2760
gaacttagag aatgcaagat ggagattgaa gacaaaaagc aggagctcct tgaaatggat 2820
caggcactta aagagagaaa ttgggaacta aagcaaagag cagctcaggt tacacatttg 2880
gatatgacta ttcgtgagca cagaggagaa atggaacaaa aaataattaa attagaaggt 2940
actctggaga aatcagaatt ggaacttaaa gaatgtaaca aacagataga aagtctgaat 3000
gacaaattac aaaatgctaa agaacagctt cgagaaaaag agtttataat gctacaaaat 3060
gaacaggaga taagtcaact gaaaaaagaa attgaaagaa cacaacaaag gatgaaagaa 3120
atggagagtg ttatgaaaga gcaagaacag tacattgcca ctcagtacaa ggaggccata 3180
gatttggggc aagaattgag gctgaccccgg gagcaggtgc agaactctca tacagaattg 3240
gcagaggctc gtcatcagca agtccaagca cagagagaaa tagaaaggct ctctagtgaa 3300
ctggaggata tgaagcaact ctctaaagag aaagatgctc atggaaacca tttagctgaa 3360
gaactggggg cttctaaagt acgtgaagct catttagaag caagaatgca agcagaaatc 3420
aagaaattgt cagcagaagt agaatctctc aaagaagctt atcatatgga gatgatttca 3480
catcaagaga accatgcaaa gtggaagatt tctgctgact ctcaaaagtc ttctgttcag 3540
caactaaacg aacagttaga gaaggcaaaa ttggaattag aagaagctca ggatactgta 3600
agcaatttgc atcaacaagt ccaagatagg aatgaagtaa ttgaagctgc aaatgaagca 3660
ttacttacta aagaatcaga attaaccaga ttacaggcca aaatttctgg acatgaaaag 3720
gcagaagaca tcaagtttct gccagcccca tttacatctc caacagaaat tatgcctgat 3780
gttcaagatc caaaatttgc taaatgtttt cacacatctt tttccaagtg tacaaaatta 3840
cgtcgctcta ttagtgccag tgatcttact ttcaaaattc atggtgatga agatctttct 3900
gaagaattac tacaggactt aaagaaaatg caattagaac agccttcaac attagaagaa 3960
agccataaga atctgactta cacccagcca gactcattta aacctctcac atataaccta 4020
gaagctgata gttctgagaa taatgacttt aacacgctta gtgggatgct aagatacata 4080
aacaaagaag taagactatt aaaaaagtct tctatgcaaa caggtgctgg tttaaatcag 4140
ggagaaaatg tgtaattcaa agaagatact gatgtgttga aaaaatggaa tttttggtac 4200
tgtgctgttt acttattata tgtagctcat acttcataga agctgttatt ttgcttttga 4260
ataaatttta tatttcaata ttttaaaaga aagcccttct aaaacttaat tatattttta 4320
aagaaaattt aaaaaaaaaa aaaaaggggg 4350

<210> 46
<211> 2959
<212> DNA

<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7505914CB1

<400> 46
ggggcgccat cttgtcttgt tcccgaagaa gtagaagcat cgaaagcgtt ggagaggtgt 60
taccggaacg gcggcgacaa gggtgttccc gaactagagt ggggcataca taatcttgct 120
gctatgcttc gaagctgtag tctgaatcaa cctaagtttt aaacagaagg tgaacctctg 180
agatagaaaa tcaagtatat tttaaaagaa gggatgtggg atcaaggagg acagccttgg 240
cagcagtggc ccttgaacca gcaacaatgg atgcagtcat tccagcacca acaggatcca 300
agccagattg attgggctgc attggcccaa gcttggattg cccaaagaga agcttcagga 360
cagcaaagca tggtagaaca accaccagga atgatgccaa atggacaaga tatgtctaca 420
atggaatctg gtccaaacaa tcatgggaat ttccaagggg attcaaactt caacagaatg 480
tggcaaccag aatggggaat gcatcagcaa cccccacacc cccctccaga tcagccatgg 540
atgccaccaa caccaggccc aatggacatt gttcctcctt ctgaagacag caacagtcag 600
gacagtgggg aatttgcccc tgacaacagg catatattta accagaacaa tcacaacttt 660
ggtggaccac ccgataattt tgcagtgggg ccagtgaacc agtttgacta tcagcatggg 720
gctgcttttg gtccaccgca aggtggattt catcctcctt attggcaacc aggacctcca 780
ggacctccag cacctcccca gaatcgaaga gaaaggccat catcattcag ggatcgtcag 840
cgttcaccta ttgcacttcc tgtgaagcag gagcctccac aaattgacgc agtaaaacgc 900
aggactcttc ccgcttggat tcgcgaaggt cttgaaaaaa tggaacgtga aaagcagaag 960
aaattggaga aagaaagaat ggaacaacaa cgttcacaat tgtccaaaaa agaaaaaaag 1020
gccacagaag atgctgaagg aggggatggc cctcgtttac ctcagagaag taaatttgat 1080
agtgatgagg aagaagaaga cactgaaaat gttgaggctg caagtagtgg gaaagtcacc 1140
agaagtccat ccccagttcc tcaagaagag cacagtgacc ctgagatgac tgaagaggag 1200
aaagagtatc aaatgatgtt gctgacaaaa atgcttctaa cagaaattct gctggatgtc 1260
acagatgaag aaatttatta cgtagccaaa gatgcacacc gcaaagcaac gaaaggtgga 1320
ctgggtggtt atggatcagg agacagtgaa gatgagagga gtgacagagg atctgagtca 1380
tctgacactg atgatgaaga attacggcat cgaatccggc aaaaacagga agctttttgg 1440
agaaaagaaa aagaacagca gctattacat gataaacaga tggaagaaga aaagcagcaa 1500
acagaaaggg ttacaaaaga gatgaatgaa tttatccata aagagcaaaa tagtttatca 1560
ctactagaag caagagaagc agacggtgat gtggttaatg aaaagaagag aactccaaat 1620
gaaaccacat cagttttaga accaaaaaaa gagcataaag aaaaagaaaa acaaggaagg 1680
agtaggtcgg gaagttctag tagtggtagt tccagtagca atagcagaac tagtagtact 1740
agtagtactg tctctagctc ttcatacagt tctagctcag gtagtagtcg tacttcttct 1800
cggtcttctt ctcctaaaag gaaaaagaga cacagtagga gtagatctcc aacaatcaaa 1860
gctagacgta gcaggagtag aagctattct cgcagaatta aaatagagag caatagggct 1920
agggtaaaga ttagagatag aaggagatct aatagaaata gcattgaaag agaaagacga 1980
cgaaatcgga gtccttcccg agagagacgt agaagtagaa gtcgctcaag ggatagacga 2040
accaatcgtg ccagtcgcag taggagtcga gataggcgta aaattgatga tcaacgtgga 2100
aatcttagtg ggaacagtca taagcataaa ggtgaggcta aagaacaaga gaggaaaaag 2160
gagaggagtc gaagtataga taaagatagg aaaaagaaag acaaagaaag ggaacgtgaa 2220
caggataaaa gaaaagagaa acaaaaaagg gaagaaaaag attttaagtt cagtagtcag 2280
gatgatagat taaaaaggaa acgagaaagt gaaagaacat tttctaggag tggttctata 2340
tctgttaaaa tcataagaca tgattctaga caggatagta agaaaagtac taccaaagat 2400
agtaaaaaac attcaggctc tgattctagt ggaaggagca gttctgagtc tccaggaagt 2460
agcaaagaaa agaaggctaa gaagcctaaa catagtcgat cgcgatccgt ggagaaatct 2520
caaaggtctg gtaagaaggc aagccgcaaa cacaagtcta agtcccgatc aaggtagtat 2580
actttttaaa gtattttgtc tgatttttaa aaaaaattga ctgaatttat tcaaagttga 2640
aagtgtcctt tctctctctc tttaataaac tcagtttggt acttgataaa taatcatagt 2700
cttaaatgtt agaaatccta tataatatta tttatttaaa attgcagatt tttaatttaa 2760
aatacatttt tatttttaaa ttttgtcttt tcccttttttt tttcagatca acaacccctc 2820
cccgtcgtaa acgctgagga atgatgtggc aagaatgcca tgatgttctt taaaaaattc 2880
catgagtttt aagggcttgt ctcaggatag aggcacattg tggctgtgta ggtgaaacag 2940

aatcttttttt tttttttaat 2959